1978

Processes In The Understanding And Improvement Of Clinical Judgments: Labels And Other Biases In Inferential Accuracy

Michael Joseph Paquin

PROCESSES IN THE UNDERSTANDING AND IMPROVEMENT OF CLINICAL JUDGMENTS: LABELS AND OTHER BIASES IN INFERENTIAL ACCURACY

by

Michael Joseph Paquin

Department of Psychology

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
July, 1978

ABSTRACT

Previous research suggests that numerous biases, particularly labels, modify clinical judgments and often have stigmatizing, or otherwise undesirable, consequences. Methodological problems such as inadequate measures and designs, confounded experimental manipulations, task meaninglessness (or irrelevance), unsatisfactory criteria for judgmental accuracy, the lack of knowledge about the critical dimensions used by judges, and the external validity of analog studies were reviewed. The approach to the present three studies with respect to these problems was outlined. In this respect a recent model of inferential accuracy was reviewed. The model is composed of two components descriptive of a judge's responses: threshold, the propensity to attribute psychopathology given limited, relevant information, and sensitivity, the awareness of actual trait inferential relations, or in other words, the judge's ability to predict a target's responses.

The first study identified five stable dimensions that judges utilized in making similarity judgments of a set of 26 psychopathology descriptors from a diverse domain representing lay labels, diagnostic categories from the second edition of the Diagnostic and Statistical

Manual (DSM II), and stimuli from a new nosology of psychopathology known as modal profile analysis. These five dimensions were interpreted as: Mental Illness, Social vs. Psychological Deviance, Depression vs. Impulse Control, Anxiety vs. Impulse Control, and Paranoid vs. Impulse Control. Results did not, overall, support the superiority of the DSM II descriptors over labels nor the beneficial influence of diagnostic qualifiers, but did suggest that information specificity, as exemplified in the modal profile stimuli, may reduce the influence of stigmatizing conceptualizations on psychopathology judgments.

The second study investigated the effects of labels ("former psychiatric patient" vs. "normal medical background"), diagnostic suggestions of symptom severity, source credibility (prestige and status of the assessment author), target behaviors, and sex of judge on inferential accuracy and other parameters of clinical judgment. The dependent variables were orthogonalized and the resulting set of seven, uncorrelated factor scores were subjected to analyses of variance. Contrary to predictions and the majority of previous studies, labels did not exert a direct, negative bias on any of the seven dimensions of judgment. In fact labels influenced prognosis scores in a positive direction; this was explained in terms of differential expectations, related to specific judgmental contexts. The psychopathological behavior patterns (targets) and diagnostic suggestions of severity respectively, were the most powerful influences on every dimension of judgment. Labels, source credibility and sex of judge, while not influential in and of themselves, were involved in numerous interactions which were interpreted as judgmental context effects of great relevance

in clinical inference and decision making.

The third study followed a similar paradigm and method of analysis as did study two. It focused on target behavior, sex of judge, informational consistency and cognitive (judgmental) focus. Again, the pattern of psychopathology represented by a target was clearly the most powerful effect. Both familiarity of the judge with the pattern, and psychopathological content of the target appeared to be involved in this target effect. The level of consistency between a label and a target did not, as hypothesized, influence judgments. Only minimal evidence was found for the effectiveness of judgmental focusing instructions intended to differentially focus judges' attention on appropriate information. Thus, it was concluded that more intensive and specific training programs are necessary in order to teach methods of attaining increased judgmental accuracy. Suggestions for improving accuracy were offered in the final chapter.

ACKNOWLEDGMENTS

As this thesis enterprise approaches completion I am humbled by my recall of the numerous sources of assistance and inspiration that have benefited me. Writing my thoughts of appreciation for all of these people, including my family, friends, peers, and teachers, is probably the most difficult part of the thesis. It would seem that there are few words which convey my truly sincere, thanks to all of you so I shall also try and do so in person.

For their helpful comments and suggestions at the proposal stage and as the thesis evolved I thank Jim Neufeld, Phil Rushton, Gus Scheid, and especially Bart Jessup who provided valuable feedback and contributed important encouragement at almost every stage. A similar note of thanks is extended to John Meyer and Erich Strasburger. Despite being in the same situation of completing a thesis, John and Erich were most helpful and supportive. I surely would not have been done at this point had they not been nearby these many months of long nights.

Finally, I am extremely greatful for and appreciative of the time consuming efforts of Douglas Jackson, my advisor. Over the past five years he has taught me so very much and has always been a source of intellectual stimulation and inspiration as well as a model of research and writing which I have tried to emulate.

This thesis is dedicated to my wife Ingrid who so unselfishly worked so hard with me at every stage. In helping with the literature review, the organization and running of the experiments, the data analysis, and the typing of the thesis, and by providing perpetual encouragement and smiles despite the demands of her own full time career, my warmest and special thanks go to her.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

Issues in clinical judgment

Interest in the clinical judgment process has enjoyed a long and exciting history. Bieri, Atkins,Briar, Leaman, Miller and Tripodi (1966) describe the history of the field of clinical judgment as a sequence of four phases. The first phase involved the use of introspective analysis in the development of psychoanalytic personality theory. Next came concerns about reliability and validity, particularly in regard to the diagnostic utilization of projective tests, an outgrowth of the first phase. Third, the work on clinical vs. statistical prediction, an extension of the validity issue, led to interest in the nature of the judgment process. The fourth phase, the development of theoretical and mathematical models of clinical judgment (e.g., information theory, the "lens" model), is a logical sequence to the expanded focus upon processes.

Recent research has covered various aspects of all four phases. In the face of growing interest in cognitive psychology, recent reviews have seen fit to re-emphasize the classical critique of introspective data, in terms of subjects' inability, in many contexts, to report accurately their cognitive processes (e.g., Nisbett & Wilson, 1977). Psychodiagnosis is again being scrutinized with respect to reliability (e.g., Spitzer & Fleiss, 1974) and validity (e.g., Rosenhan, 1973; Schact & Nathan, 1977). Increasing use of computerized test scoring and interpretation is an efficient extention of statistical prediction, and the recent work of Jackson (Jackson, 1972; Reed & Jackson, 1975)

on inferential accuracy exemplifies work on new models of clinical judgment.

This thesis represents what might be considered a fifth phase of clinical judgment research, an extention of the validity issue in conjunction with the concern with models of judgment. A series of three studies is presented having as their purpose the investigation and ultimately the improvement of accuracy in clinical judgment, using Jackson's model of inferential accuracy.

Bieri et al. (1966) have outlined four categories of variables, in discrimination judgments of the kind involved in psychodiagnosis, which are relevant here.

> "These include (1) stimulus variables, or the nature of the input, (2) output or response variables, including the nature of the judgment task, (3) characteristics of the judge, who must function as a transmitter of the information, and (4) situational variables, which include the kinds of restraints imposed upon the setting within which the judgment occurs, as well as the expected outcome or consequence of the judgment," (p.8).

The author will, in the final chapter, interpret some of the major findings in terms of "context" variables, which can be conceptualized as a superordinate construct involving the interaction of these input, output, and situational variables. Characteristics of the judge such as sex and training are also considered.

The Psychodiagnostic Process

Psychodiagnoses, prognoses, and other related aspects of decision making about patients are some of the most important applications of clinical judgments. It is little wonder that much has been written

about the nature, reliability and validity of diagnostic enterprises.

A basic problem in evaluating the utility of diagnostic practices is that they have often served more than one purpose. For example, classification does have some administrative and legal utility with respect to the assignment of patients to treatments or determining a legal status of sanity. Aside from ideological criticisms opposed to the general notion of classifying people, diagnosis is also criticized for its major use, namely, the attempt to relate treatment of psychopathology to the description of processes underlying observed symptoms and their etiology. Many years ago Zigler and Phillips (1961) wrote:

> "The problem of validity lies at the heart of the confusion which surrounds psychiatric diagnosis. When the present diagnostic schema is assailed, the common complaint is that class membership conveys little information beyond the gross symptomatology of the patient and contributes little to the solution of the pressing problems of etiology (cause), treatment procedures, prognosis, etc. The criticism that class membership does not predict these important aspects of a disorder appears to be a legitimate one. This does not mean that the present system has no validity," (p.612).

More recent reviews suggest that the situation, in general, has not improved substantially. Costello (1970) argues that "present classification systems are inadequate primarily because they are premature," (p.1); he advances nine reasons for this conclusion, the most important of which are that interjudge agreement is low, diagnoses are unstable and show little evidence of concurrent validity, and that there does not seem to be a consistent relationship between diagnosis and treatment. Many of the same criticisms and others are put forth by other authors (e.g., Arthur, 1969; Chapman & Chapman, 1967, 1969; Hunt, Schwartz & Walker, 1965; Schectman, 1967; Zubin, 1967).

Rosenhan (1973), in what is perhaps one of the most controversial studies of the problem, concludes that current diagnostic procedures cannot even distinguish the sane from the insane. Even the most recent comprehensive diagnostic systems, while substantially superior to their predecessors, are subject to criticism (Schacht & Nathan, 1977).

There are a number of undesirable consequences of such fallible diagnoses for the patient. Aside from the possibility that an unreliable and invalid diagnosis may lead to unsuitable treatment, the attempt to predict a target patient's behavior, prognosis, and proper therapy may be biased by the use of attributes imputed to the group supposedly represented by a diagnostic category. This has often been called "stereotyping." Furthermore, the individual may experience various forms of social rejection or alienation, i.e., may be stigmatized by the label associated with the stereotype. It is possible that labels also influence professional judgments about the person. What is perhaps more serious is that possibility that these judgmental labeling effects may become self fulfilling prophecies; our erroneous (stereotypical) expectations may lead to behavioral changes that tend to result in the labelled person responding in accord with these expectations. Such hypotheses have received substantial support from research on the sociology of deviance, in the societal reaction and labeling theory tradition (e.g., Kirk, 1974, 1975; Sarbin, 1967; Sarbin & Mancuso, 1972; Scheff, 1974, 1975), a tradition which is highly congruent with recent psychological approaches to abnormal behavior (e.g., Ullman & Krasner, 1969). It should be noted however, that this perspective has not gone unchallenged (e.g., Davis, 1972; Gove, 1970a, b, 1973, 1974; Murphy, 1976). As shall be clarified later, the issue

is far from resolved and poses numerous research possibilities.

Before proceeding to a review of the consequences of biases in clinical judgments it is instructive to conclude with a powerful statement, worthy of serious scientific scrutiny, by Rosenhan (1973, p. 257):

> "Whenever the ratio of what is known to what needs to be known approaches zero, we tend to invent "knowlege" and assume that we understand more than we actually do. We seem unable to acknowlegde that we simply don't know. The needs for diagnosis and remediation of behavioral and emotional problems are enormous. But rather than acknowledge that we are just embarking on understanding, we continue to label patients "schizophrenic," "manic-depressive," and "insane," as if in those words we had captured the essence of understanding. The facts of the matter are that we have known for a long time that diagnoses are often not useful or reliable, but we have nevertheless continued to use them. We now know that we cannot distinguish insanity from sanity. It is depressing to consider how that information will be used."

Consequences of Biases: Overview

A vast number of factors (e.g., diagnostic suggestion; prestige; information and impression availability, perceived relevance, context, consistency and extremity; patient and therapist characteristics such as sex and social class; training and experience; physical and psychological characteristics such as appearance, past history, verbal ability, insight, attitude and interaction styles) have been investigated as potential sources of undesirable bias in clinical judgments. Kahneman and Tversky (1973) and Tversky and Kahneman (1974) for example, review and evaluate a variety of such biases relevant to judgmental predictions in general, and conclude that statistical or probability type errors and judgmental "heuristics," such as information representativeness are

indeed prevalent. Similarily, Abramowitz and Dokecki (1977) in an excellent review, evaluate evidence for a number of patient characteristics related to what might be called the "latent sociopolitical functions of mental health practice" (e.g., race, sex, class, values). They conclude that although evidence for some of these variables (e.g., social class) supports the presence of bias in judgments,

> "extant data suggest that clinical judgment bias is more circumscribed than critics of the mental health establishment have forecast. The scattered evidence of this worrisome phenomenon is nonetheless sufficient to warrant continued study and professional vigilance," (p. 473).

The single most studied bias is type of label. The magnitude of interest in labels is probably due to their direct link with diagnosis as well as the controversy generated by the labeling theory of deviance.

Although it can be argued that such labels are useful because they make heavy information processing manageable, the point still remains that in most interview situations labels have the deleterious effect of obscuring an objective evaluation. As Langer and Abelson (1974) point out:

> "In practical terms, the labeling bias may have unfortunate consequences whatever the specific details of its operation. Once an individual enters a therapist's office for consultation, he has labeled himself "patient." From the very start of the session, the orientation of the conversation may be quite negative. The patient discusses all the negative things he said, did, thought and felt. The therapist then discusses or thinks about what is wrong with the patient's behavior, cognitions and feelings. The therapist's negative expectations in turn may affect the patient's view of his own difficulties, thereby possibly locking interaction into a self fulfilling, gloomy prophecy," (p. 9).

Recent work on the psychology of prediction and other clinical judgment processes suggest that a number of cognitive biases that derive from the operation of judgmental heuristics such as "representativeness,"

are involved (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). That is judgments are often made on the basis of which outcome or decision appears most representative of, or similar to, the evidence. Labels perform such a function, often summarizing and simplifying cumbersome quantities of evidence. This sometimes is appropriate, but as Kahneman and Tversky (1973) demonstrate, this representative heuristic often ignores the base rate or probability of outcomes and the reliability of the evidence. In such cases and when judgments are anchored to previous, often unrepresentative experiences, expectations, or labels, accuracy suffers. This deterioration in accuracy happens despite the fact that such anchors enhance confidence in the judgments made (Kahneman & Tversky, 1973; Oskamp, 1965).

Alternatively, there is evidence in the literature that in many situations labels do not bias judgments nor do they stigmatize patients as much as has been claimed. In this regard a label versus behavior issue is beginning to emerge. The discussion now turns to evidence regarding the existence of labeling effects and the relative importance of labels, as opposed to the target's observed behavior, on clinical judgments.

Labeling Effects: Review of the Evidence

In an important study on processes and outcome in therapy, Staples, Sloane, Whipple, Cristol, and Yorkston (1976) suggest that the characteristics of patients may be more relevant to improvements than specific therapist interventions. Labels derived from prior diagnosis and previous history may well be one such characteristic. Some studies

have in fact demonstrated that labels such as "ex-mental patient" do have a dramatic, stigmatizing effect on important aspects of the lives of the recipients of such labels (e.g., Farina & Felner, 1973; Farina, Felner, & Boudreau, 1973; Farina & Ring, 1965; Kirk, 1974, 1975; Loeber & Weisman, 1975; Page, 1977; Riess, 1977; Wyatt, Reardon, & Bass, 1977). These stigmatizing consequences involve employment opportunities, attainment of accommodation, social acceptance and related concerns of importance. Alternatively, there is literature which claims that the mentally ill and people who were formerly labeled as such are not victims of stigma and rejection in such aspects of daily living (Crocetti, Spiro, & Siassi, 1974; Gove, 1970a,b, 1974; Page, 1974).

The effects of labels also reveals a similar, conflicting pattern of results on judgments of patients, even though the weight of the evidence to date does support the presence of labeling bias. Langer and Abelson (1974), for example, in a frequently cited study, showed that clinicians with a traditional psychoanalytic background, judged a taped interview of a target as significantly less adjusted when he was labeled "patient" as opposed to being described as a "job applicant." Snyder (1977) conducted further analyses of Langer and Abelson's data with respect to the clinicians' attributions of the locus of the targets' problem. Snyder concluded that:

> "... a clinician's training can also have an important effect on where the clinician 'sees' the interviewee's problem to lie - especially if that interviewee is labeled a patient. More specifically, the psychodynamically trained clinician 'located' the patient problem in the person, and the behaviorally oriented clinician 'located' the patient problem in the situation," (p. 102).

One important criticism of the Langer and Abelson study is offered by Farber (1975) who parenthetically notes that "of course, we do not know in the absence of a definitive external criterion, whether one or both groups of clinicians were in error..." (p. 606-607). A recent study by Paquin and Jackson (1977) which did employ a definitive criterion of accuracy (i.e., sensitivity) derived from a recent model of inferential accuracy (Jackson, 1972; Reed & Jackson, 1975), reached the same conclusion as Langer and Abelson. When subjects judged written descriptions of actual patient types, the presence of labels such as "psychiatric patient," "drug addict," "ex-convict," and "alcoholic" significantly increased their propensity to attribute psychopathology to the targets; this increased tendency to ascribe pathology was a source of judgmental error. Judges who received the labeled descriptions were less accurate in predicting, i.e., were less sensitive to, the described person's responses for three of the four targets. Similarily, a number of other studies have varied the presence of labeling information within the context of an otherwise standard stimulus and have observed a detrimental bias due to labels (e.g., Calhoun, Pierce, Walters, & Dawes, 1974; Del Gaudio, Stein, Ansley, & Carpenter, 1976; Lee & Temerlin, 1970; Reade & Wertheimer, 1976; Snyder, Shenkel, & Schmidt, 1976; Sushinsky & Wener, 1975).

Studies concerned with this same paradigm have also produced or noted negative effects (cf. Billingsley, 1977; Crocetti et al. 1974; Gove & Fain, 1973; Kirk, 1974, 1976; Page, 1974). This failure to find a labeling bias has formed the basis for the argument that it is the described behavior of the target rather than labels, or other biasing information, which is the major determinant of clinical judgments. The

weight of evidence, limited as it is, still appears to favor the presence of labeling effects. Evidence from different methodological approaches, not reviewed here, also appear to support this tentative conclusion. However, it is argued that final resolution of the label vs. behavior debate is premature, having been obscured by a number of methodological problems across many of these studies. A brief outline of some of these problems and the approach of this thesis is discussed next.

Methodological Issues

In their review of value bias in clinical judgment, Abramowitz and Dokecki (1977) note that four types of studies, variants of in situ correlational versus analog studies involving patient and or clinician (judge) variables, have been employed. The analog approach, where patient and/or judge characteristics are systematically varied in otherwise standardized stimulus materials, is a preferred approach because of its greater internal validity. This is the approach followed by the studies of clinical judgments, just reviewed. These studies, depending upon their level of similarity to complex, actual clinical equivalents, vary in their degree of external validity, or generalizability. One criticism levied against such analog studies is the relatively limited information and involvement permitted the judge. Similarily, it has been said of many judgment tasks, that the investigator had failed to provide the judge with a meaningful task. For example, tasks are often not relevant to clincal decision making, and information sources (e.g., projective or MMPI test results) may or may not relate to known empirical criteria or have substantive relationships (to theory or

experience) of the kind in the task itself.

Abramowitz and Dokecki (1977) also note that most studies have employed global ratings, such as "personality impressions, and diagnostic, and prognostic judgments representative of traditional psychiatric practice," (p. 463). Similarily in the studies reviewed above, ratings of psychopathological severity, adjustment, and desirability are common. Earlier it was noted that the Langer and Abelson (1974) study failed to provide an external criterion of accuracy. This is not, of itself, a major shortcoming since many studies are concerned primarily with the relative effects of some variable (e.g., a diagnostic label) on judgments. However, the virtual absence in the literature of what Farber (1975) has termed "definitive external criteria" of accuracy is indeed a shortcoming. It may be that judges in general are fairly accurate, as might be predicted by implicit personality theory, and that the observed differences in accuracy in judges as a function of experimental manipulations may, in an absolute sense, be rather small.

A number of problems arise across studies due to the number of isolated components of a clinical judgment analog which are chosen for study. The investigation of a small number of factors only amplifies the discrepancy between the analog situation and the clinical counterpart, and often contributes to task meaninglessness by depriving the judge of important and typical sources of information. For example, in order to answer the label vs. behavior debate the research must systematically and orthogonally vary labels and behavior in the stimulus materials. But more important, these ought to be varied

together with other important aspects of the clinical situation if generalizable results which advance our knowledge of clinical judgments in their real-world contexts, are to be obtained. To be sure, there is a dilemna inherent in such complex research efforts, namely, the confounding (e.g., of labels with behavior and suggestion effects, or, prestige factors with suggestion and labels etc.) of these multi-dimensional judgmental contexts. However, these are the kinds of experimental manipulations, with their corresponding independent measures, which appear to be needed at this stage.

The tendency to employ only one or two variables is also problematical with respect to attempts to compare findings across studies. The tendency of investigators to use and/or invent their own preferred dependent measures only exaggerates this problem. Consider, for example, the range of possible types of dependent measures (usually global ratings) noted above. These can also, of course, vary in terms of specific content thereby producing a practically infinite range of measures. Since the effects of biases are likely to vary, at least across the major dimensions of clinical judgment, it is little wonder that it is difficult to synthesize the results of the literature.

Few studies have even attempted to identify the major dimensions which different types of judges utilize in specific content domains. Thus, with but a few exceptions (e.g., Chan & Jackson, 1978; Cox, Costanzo, & Coie, 1976; Messick & Jackson, 1972), little is published which helps to illuminate major, judgmental dimensions of psychopathology in clinicians and popular conceptions of "mental illness" in general. Until stable dimensions of judged psychopathology can be

identified it will be extremely difficult to conduct conclusive evaluations of the differential effects of various biasing and general factors on specific dimensions of clinical judgments.

Approach to this Thesis

This series of studies attempts to address a number of these problems. One of the concerns of the first study, for example, is the identification of the major dimensions underlying judgments of psychopathology across three categories of psychopathology descriptors. One of the outcomes of this will be the determination of the level of consistency or similarity between labels, i.e., lay conceptions of mental illness, and empirically identified behavioral exemplars of psychopathology. This scaling will be utilized in the first study to generate combinations of labels and target behaviors which differ in their mutual, perceived consistency.

The second and third studies, which investigate labeling and other biases in clinical judgment, also address a number of methodological issues. Both employ a model of inferential accuracy and therefore utilize an external criterion of accuracy against which these biases can be evaluated. The model of inferential accuracy is described in full by its author (Jackson, 1972) and evaluated in Reed and Jackson (1975).

> "Inferential accuracy is defined in terms of a person's ability, given limited information about a target person (patient), to judge correctly other pertinent characteristics about that person and to identify behavioral exemplars as part of a pattern of behavioral consistencies," (Reed & Jackson, 1975, p. 475).

Briefly, the model defines two processes: Threshold is the generalized tendency to attribute psychopathology to a target and is operationalized as the judge's mean (rating) response on the items to be predicted for each target. Sensitivity can be defined as an accurate awareness of trait inferential relationships and is operationalized here as the correlation between the judge's predictions of the targets' responses and the actual scores of the profile for the type of patient, or person, represented by the target. Thus, sensitivity is a "definitive external criterion," representative of the judge's ability to predict the responses of actual patient types. Actually, a number of criteria for sensitivity are possible: (1) a modal profile (2) a profile exemplar and (3) a group consensus value. Reed (1976) has shown that these three criteria are highly correlated across targets defined by both pathological and non-pathological traits. The criteria vary somewhat, however, in their stability. The advantage of the criterion employed in the present studies, the modal profile, is that it possesses considerable stability, i.e., it generalizes across populations of patients (Skinner, Reed, & Jackson, 1976) and yet directly reflects the responses of actual patients.

This brings up another advantage of the present approach. The targets which subjects judged, being based on actual patient responses, were, strictly speaking, neither contrived nor hypothetical, as has been the case in many studies. The targets were derived from studies of large samples of respondents to the Differential Personality Inventory (DPI) of Jackson and Messick (1971) according to a procedure called modal profile analysis (cf. Skinner, Jackson, & Hoffman, 1974; Skinner et al., 1976). The resulting modal profiles represent empirically identified, mutually distinct, types of psychopathology which demonstrate

a relatively high generality and classification efficiency across diverse patient populations.

Because these profiles and the item set upon which judges predict target responses were both derived from a construct oriented inventory (the DPI) the judgment task in the second and third studies facilitate the use of theoretically meaningful links among the judged stimuli thereby potentiating greater accuracy than has been found in earlier studies. As Reed and Jackson (1975, p. 481) put it

> "... the use of explicit dimensions of psychopathology is one crucial aspect that distinguishes this approach from many of its predecessors. The judged targets represent empirically identified types based on clearly defined dimensions of behavior. Evidence has been presented (Jackson & Carlson, 1973) for the convergent and discriminant validity of measures of these dimensions of behavior. In addition, the items used were chosen from a set of behavioral exemplars devised to represent a wide situational range. As a result of these factors, the relations among stimuli are simply more lawful and their convergence with relevant external behaviors greater than has been the case for stimuli used in previous clinical judgment research. In short, the task is more meaningful than the majority of other experimental clinical judgment tasks."

The present studies exploit this task meaningfulness in order to study selected judgmental biases. In this regard the number of such information sources serves to enhance task meaningfulness by providing subjects with a judgmental context and task which, more than many studies, approximates the rich information base in actual clinical decision making. Despite the relatively large number of variables in study two and three, confounding of variables is avoided by the systematic, orthogonal, manipulation of the key bias variables.

CHAPTER TWO

The Dimensionality and clinical referents of labels and psychiatric nomenclature of psychopathology

Overview

This series of studies explores processes underlying the effects of labels and other forms of semantic and conceptual information upon clinical judgment. Taking as a point of departure a recently published study (Paquin & Jackson, 1977), they explore the effects on conceptions of psychopathology of relevant, as well as pejorative prior information.

The first of these studies attempts to identify, in this context of labeling effects, the underlying dimensions of psychopathology judgments. In other words, what are the major constructs or conceptualizations or perspectives that a judge uses when processing behavioral and trait information. This is critical to an understanding of the mechanisms and processes by which labels and other information sources exert their effects on clinical judgments.

To date the literature has been inadequate in articulating the nature and number of important dimensions of perceived psychopathology. There is a considerable body of research which attempts to explain observed dimensions of psychopathology including efforts to identify reliable patterns or types of disorders (eg., Hoffman & Jackson, 1976; Hoffman, Jackson, & Skinner, 1975; Morf, Syrotuik, & Krznaric, 1977; Skinner, Jackson, & Hoffman, 1974; Skinner, Reed, & Jackson, 1976).

In fact this study incorporates several "types" or profiles from one such recent approach to classification which has shown considerable promise (Skinner et al., 1976).

However, dimensions based upon observed patterns of behavior or test responses may not correspond to the dimensions relevant to a judge's (e.g., psychiatrist) perceptions and conceptualization of that psychopathology. It is the latter, of course, which bears directly upon the clinical judgment process and herein lies the deficiency in the literature.

In contrast, judgmental dimensions of personality have been reasonably well studied and there is in fact good evidence that perceived dimensions of personality correspond to observed dimensions. This is best illustrated by studies showing that the implicit personality theory of judges is, in general, accurate by virtue of the correspondence of perceived vs. actual trait covariations (Lay & Jackson, 1969; Stricker, Jacobs, & Kogan, 1974).

Chan and Jackson (1978) provide one of the few recent attempts to study these matters in the context of psychopathology. Fitzgibbons and Shearn (1972) provide some data on the conceptions of clinicians with respect to one disorder, namely, schizophrenia. Messick and Jackson (1972) uncovered 13 dimensions of psychopathology based on desirability judgments of MMPI items. The present investigation continues this line of study by attempting to clarify further the dimensionality of psychopathology judgments using a diverse domain of psychopathology descriptors.

Three categories of such descriptors varying in their informational specificity and other characteristics are explored here. The first category is lay labels, i.e. general, everyday terms that are

used to describe socially and psychologically salient behavior problems, e.g., "mentally ill." This type of descriptor has been frequently examined in studies of attribution and stigma (cf. Aloia, 1975; Cauthen, Robinson, & Kraus, 1975; Farina & Felner, 1973; Farina & Ring, 1965; Gove & Fain, 1973; Kirk, 1976).

The second category of descriptors to be studied are taken from the major nosological system for psychopathology, namely, the Diagnostic and Statistical Manual (DSM) of Mental Disorders II (hereafter, DSM II). The DSM II is published by the American Psychiatric Association (1968) and covers the entire range of psychopathology in some detail. More importantly, it is probably the most widely accepted and utilized classification scheme for psychopathology and consequently ten of the stimuli studied in this investigation are taken from the DSM II.

The third category of descriptors represent a modern approach to the classification of psychopathology based upon psychometric principles. Empirically identified and cross-validated patient types or "modal profiles" are used to generate brief summary descriptions focusing on specific behavioral examplars of the psychopathology in question. There are eight such discrete profiles.

The original goal of this study was the scaling of similarity or consistency of the stimuli among the three categories of descriptors. This was necessary for the experimental manipulations of the second and third studies in this thesis. However, it became clear that a closer look at the dimensionality of the psychopathology descriptors would be of great value for its own sake in view of the lack of research to date

on this topic.

Purpose

The purpose of this study is to examine and map out the underlying dimensions of various concepts and descriptors of psychopathology. It is hypothesized that only a few dimensions are necessary to explain the underlying structure of such concepts. It is further expected that this simple dimensionality holds irrespective of whether vague labels of psychopathology based upon lay personality theory or more specific nosological terms such as those defined by the DSM II are used. Labels are likely to be proximally located to diagnostic categories in multidimensional space, suggesting that both may function in the same way in clinical judgments.

A number of corollaries can be formulated for the primary hypothesis that a few major dimensions can explain clinical judgments of psychopathology. If the most important of these dimensions reflects a gross, overgeneralized and denotatively pejorative, socially undesirable conceptualization, then stimuli (and patients) which are assigned vague labels are more likely to be explained and evaluated by such a conceptualization than are stimuli with more specific, less "stereotyped" labels. In general, it is hypothesized that the informational specificity of psychopathology stimuli is inversely related to broad, evaluatively based conceptualizations of psychopathology; it is predicted that the latter will account for most of the variance in clinical judgments.

Furthermore, given the powerful influence of such a conceptualization it is predicted that moderating qualifications to diagnostic evaluations will be ineffective in differentiating clinical judgments. Some evidence for this is noted by Rosenhan (1973) in his classic and controversial investigation: he observes the consequences to putatively schizophrenic "pseudopatients" with respect to the use of the term "in remission," in the diagnosis at discharge. The latter as well as the qualifiers "not presently psychotic" and "mild," all integral qualifiers in the DSM II, are investigated in this study. It is hypothesized that judgments of stimuli will not be significantly altered by these three qualifiers nor by adjectives connoting a prior status e.g., "former" psychiatric patient.

A Dimensional Approach to Psychopathological Concepts

The question addressed by this study is not unlike those addressed in studies investigating various aspects of implicit personality theory or trait-inferential relationships (e.g., Jackson, 1969; Rosenberg & Sedlak, 1972a,b). It is primarily the content which differs here; we are interested in establishing the nature of the psychological structure of the three sets of psychopathology descriptors. This can be accomplished in a number of ways. One could, for example, ask judges to rate a set of labels and diagnostic categories on bipolar, semantic differential scales and then analyse the results comparatively.

The difficulty with this approach is that the bipolar adjectives are not psychologically opposite for all people, or even if so, not necessarily to the same extent when comparing across the different adjective pairs. It may therefore be difficult to interpret the resulting structure. Similarly, one could obtain trait ratings of the labels and diagnostic categories and factor analyse the correlations among these ratings to indirectly construct a psychological structure for these concepts.

Alternatively, multidimensional scaling analyses, which generate a geometric representation of the underlying structure or pattern of relationships among a set of judged stimuli, can be employed. Such scaling methods are particularly appropriate when there are substantially fewer dimensions than concepts and when little is known about the nature of the most important underlying dimensions. As Jackson, Messick and Solley (1957) pointed out such methods avoid the necessity of predetermining the number and nature of relevant dimensions which are largely unknown. These scaling methods have been used in a number of studies of judgmental processes (cf. Chan, 1975; Jackson, 1962, 1969; Jackson & Messick, 1963; Partington & Jackson, 1968; Porporino, 1974; Rosenberg & Sedlak, 1972a,b; Walters & Jackson, 1966) which point to the utility of multidimensional scaling in trait inference investigations of perceived personality and psychopathology.

The metric scaling approach to be used in this study analyzes the matrix of similarity judgments formed from all possible paired combinations of the stimuli (i.e. psychopathology descriptors) to be judged. Values are obtained which represent an estimation of the distance between any two stimuli. Highly similar stimuli are represented by (distance) points which are close together in the "space" representing

the judgments. The matrix of all these values generates a few dimensions which map a geometric configuration of the stimuli in terms of interpoint distances interpreted as the large components of psychological relatedness.

In summary, the purpose of this study is to discover the nature of these basic dimensions. To facilitate the interpretation of these dimensions, the relationship between these dimensions to semantic differential ratings and to prognostic clinical judgments will also be examined. The major focus will be on the similarity of three types of conceptual representations of psychopathology: (a) empirically defined personality types, (b) various labels, and (c) the DSM II psychiatric nomenclature.

METHOD

Subjects

The 46 "Judges" (subjects) were primarily graduate students in clinical psychology who had completed at least one year of their program including an internship (n=27), and undergraduates who were completing an advanced course in abnormal psychology (n=19). Selecting judges in this way at least insures a degree of familiarity with psychopathology and diagnostic systems. The undergraduate

students were offered $6.00 as compensation for their time; the average time needed to complete the task was about three hours.

Procedure

The judges were asked to rate the similarity of the experimental stimuli on nine point scales ranging from very similar to very dissimilar. They were instructed simply to examine each stimulus pair briefly and check off on the scale their similarity judgment according to their first impressions.

Next, the subjects rated each of the experimental stimuli individually. These unidimensional ratings included semantic differential scales and selected general, clinical judgments concerning each of the stimuli (see Appendix A). For example, they were asked to indicate their estimate of the following: (i) psychopathological severity, (ii) the need for institutionalized care, (iii) social (un)desirability, (iv) their estimate of the base rate or frequency of occurrence, of the stimuli in a random sample. They also rated their confidence in the judgments they made.

Experimental Stimuli

The first set of stimuli are seven brief paragraph descriptions

of people representing different psychopathological types, known as modal profile types, based upon scale scores from the DPI i.e., the Differential Personality Inventory (see Appendix D). The eighth stimulus was a description of a relatively normal individual. These "types" were empirically identified and validated by a procedure known as multiple classification or "modal profile" analysis (Skinner et al., 1974; Skinner et al., 1976). A modal profile is defined as a hypothetical pattern that is characteristic of a certain subset of patients in the population. "Each modal profile represents a unique pattern of DPI scale scores that was characteristic of frequently occurring profiles in the derivation sample .... These modal profiles are based on a dimensional classification model and essentially describe a parsimonious eight-dimensional space into which individual profiles may be projected using a factor extension rationale ...," (Skinner et al., 1974, p. 112).

A least squares estimate of the profiles is derived through a sequential analysis based on a generalized canonical correlation procedure. An important facet of this strategy is the assessement of the independent contribution of elevation, scatter, and shape parameters in defining profile similarity. Even more important is the evidence for the generalizability of these modal profiles across diverse populations and their relatively high classification efficiency (Skinner et al., 1976).

In summary, there is considerable evidence for the psychometric viability of the taxonomy based on the modal profile approach. Furthermore, because it is based upon a construct-oriented inventory (DPI) these profiles facilitate the use of meaningful links among the

stimuli in this study. This overcomes the problem of "meaningfulness" and "fairness" that has plagued many clinical judgment tasks.

The defining characteristics of the seven modal profiles and the written descriptions based upon these profiles can be found in Appendix A. This approach to describing psychopathology in the context of the clinical judgment process is exemplified in previous studies (Paquin, & Jackson, 1977; Reed & Jackson, 1975).

The second set of stimuli include various lay labels of psychopathology, some of which have been investigated in the previously cited studies of attribution and stigma. These include "former psychiatric patient," "alcoholic," "ex-convict," "psychiatric hospital patient," "drug addict," "legally insane," "mentally disturbed," and "nervous". The latter is included as a contrasting stimulus, connoting, at most, mild psychopathology.

Finally, a set of diagnostic categories commonly found in psychiatric nomenclature were examined. While this does not represent a random sample of the nomenclature, it is adequate for the specific exploratory purposes of this study. These 10 stimuli were taken directly from the DSM II of the American Psychiatric Association (1968). Originally a larger set of stimuli was planned however, this was reduced to the present set for pragmatic reasons, primarily involving the length of time required to complete the judgments of all the stimulus pairs. The major focus was on the frequently occurring disorders such as anxiety, depression and schizophrenia. In addition separate stimuli were defined on the basis of supplementary, qualifying phrases utilized by the DSM II, i.e., "not presently psychotic," "mild" and "in remission." The latter, for example, was

a specific issue in the controversy over Rosenhan's paper "On being sane in insane places" (cf. Farber, 1975; Rosenhan, 1975; Spitzer, 1975; Weiner, 1975). Thus, eight of the diagnostic stimuli are: "Schizophrenia, simple type,""Schizophrenia, simple type, in remission," "Schizophrenia, paranoid type," "Schizophrenia, paranoid type, not presently psychotic," "Anxiety neurosis," "Depressive neurosis," "Manic depressive illness, depressed type," "Manic-depressive illness, depressed type, mild." In addition, two more frequent but less serious categories were included to test for the ability of the judges to make differential judgments about psychopathology. These are: "Marital maladjustment" and "Alcoholism, episodic excessive drinking." Thus, there are a total of 10 stimuli from the DSM II which are taken to represent the current psychiatric diagnostic system. There are 26 stimuli overall in the study: the ten DSM II stimuli, the eight descriptions based on the modal profiles and eight labels, or lay conceptualizations of psychopathology.

Method of Analysis

The total number of paired judgments of these 26 stimuli is 325. These 325 pairs presented in quasi-random in order to minimize order effects and the time required to re-order the data, were analyzed by a principal components, successive intervals model of multidimensional scaling (Diederich, Messick & Tucker, 1957; Messick & Abelson, 1956; Torgerson, 1958).

This model is composed of three parts. First cumulative proportions of the raw data are converted to z-scores corresponding to

the area of the normal curve represented by each cumulative proportion. Next, in the "successive intervals scaling," category boundaries, dispersions and scale values are computed for each stimulus. Finally, the dimensionality, stimulus projections and eigenvectors of the relative distance matrix are computed from the n(n-1)/2 inter-stimulus distances derived from the successive interval scaling. The obtained dimensions are then rotated to the varimax criterion.

The analysis of the unidimensional stimulus ratings of the second part of the task began by computing the means (across judges) for each stimulus on each rating question. The purpose here is to identify general factors rather than individual differences. The 26 X 10 matrix (stimuli X ratings) was then factored (major product moment solution) using a singular value decomposition procedure following the Eckart-Young theorem (Horst, 1965).

The 10 X 26 (ratings X stimuli) matrix of this analysis (minor product moment solution) was also examined in order to investigate the structure underlying the rating questions and relate this to the interpretations of the other analyses. Finally, the same analyses were repeated using a principal factor as well as a principle components solution for the (26 X 26) correlation matrix for the stimuli. The resulting structures were virtually identical for all analyses, a finding which might be expected from factor theory (Horst, 1965), and so they are not reported. The same is true for the results reported in Table 4, the factor analysis of the (10 X 10) correlation matrix for the rating questions. Canonical correlations were used to assess the similarity of entire matrices. This involves a linear combination of the factors in each matrix such that the correlation between them is maximized (comparisons of individual factors are not appropriate).

RESULTS

Table 1 depicts the five-dimensional solution of the multidimensional successive intervals scaling after orthogonal rotation to a varimax criterion. The eigenvalues for the first ten dimensions were: 27.43, 17.07, 14.35, 8.09, 5.54, 4.84, 4.18, 3.34, 2.95, and 1.83 respectively. It is apparent that the eigenvalues drop off quickly in magnitude after three dimensions. However, in order to provide a conservative estimate of the dimensionality, five dimensions were rotated and are interpreted below. These five dimensions account for 95.4% of the total variance.

The size of each eigenvalue (or latent root) may be interpreted as the relative amount of variance accounted for by the dimension. In general, the first few dimensions account for most of the reliable variance. The determination of the appropriate dimensionality is often facilitated by plotting the relative sizes of the eigenvalue for each dimension. Figure 1 depicts such a plot. In the present case it would appear accurate to define the first three, or, to be safe, the first five dimensions as sufficient to represent the judgmental dimensions of psychopathology of this group of subjects.

Stability of the Dimensions

The stability across samples of the obtained dimensions in a multidimensional scaling solution is an important concern of such analyses. One way of assessing stability is to divide the total sample into random halves and compute correlations between the respective

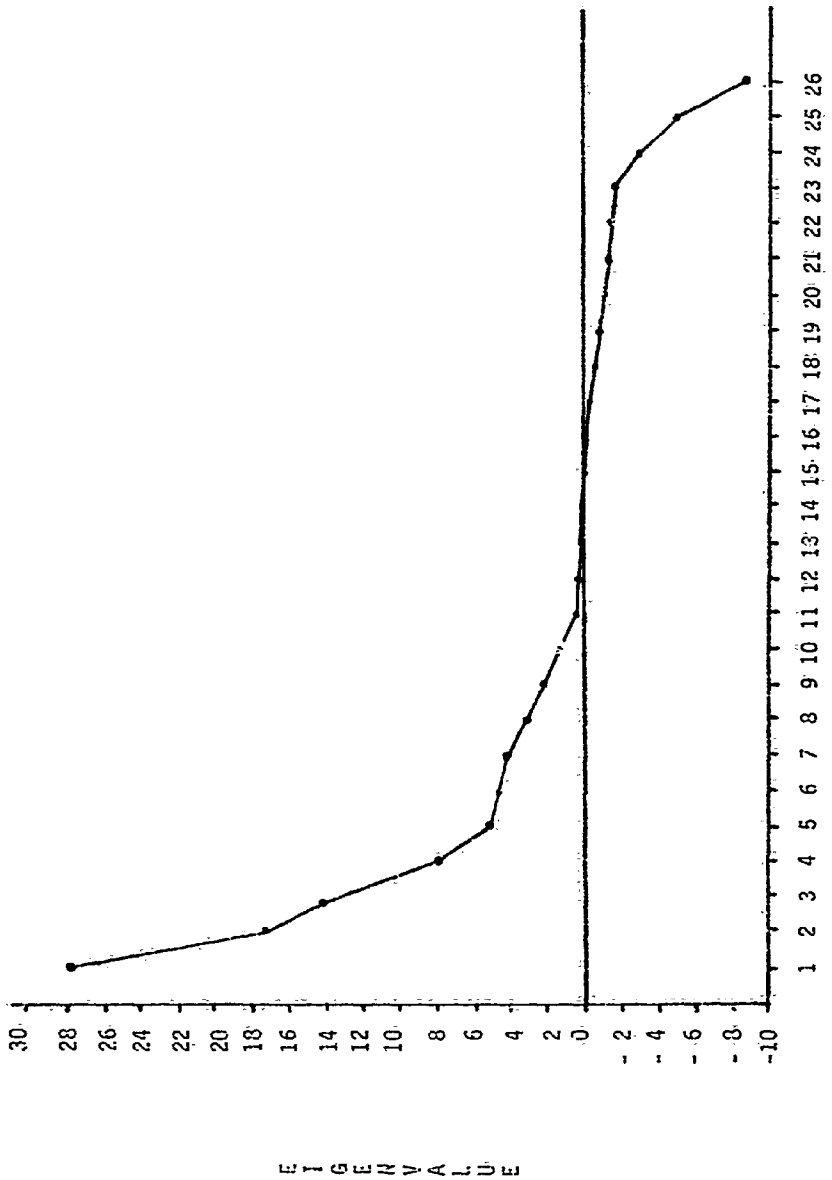
Table 1

The Rotated Multidimensional Scaling Configuration

of Psychopathology Descriptors

	Stimulus	Dimension				
		I	II	III	IV	V
1	Former psychiatric patient	-.70	-.63	-.07	-.29	-.05
2	Alcoholic	-.41	1.12	.25	.10	-.55
3	Ex-convict	.42	1.16	-.92	-.81	.04
4	Psychiatric hospital patient	-.65	-.24	.21	-.14	.35
5	Drug addict	.14	1.14	-.09	.20	-.76
6	Insane (person)	-2.37	.10	-.15	-.11	.45
7	Mentally Disturbed (person)	-.56	-.28	.12	-.37	.49
8	Nervous (person)	.44	.10	-.46	1.27	.13
9	Schizophrenia, simple type	-1.25	-.85	-.27	-.25	-.20
10	Schizophrenia, simple type in remission	-.15	-1.29	-.60	-.14	-.45
11	Schizophrenia, paranoid type	-.55	-.08	-.68	.40	1.41
12	Schizophrenia, paranoid type not presently psychotic	-.20	-.33	-.65	.13	1.00
13	Anxiety Neurosis	.13	.11	.16	1.53	-.11
14	Depressive Neurosis	.02	.27	1.48	.39	-.16
15	Manic-Depressive illness, depressed type	-.37	-.07	1.68	-.23	-.07
16	Manic-Depressive illness, depressed type, mild	.05	-.14	1.47	-.12	-.11
17	Marital maladjustment	-.30	.73	-.09	-.09	-.17
18	Alcoholism episodic excessive drinking	.54	1.16	.21	.07	-.41
19	Alan Hill	.06	-.04	-.64	-.42	1.45
20	Jack Cole	-.14	.93	-1.11	-1.66	-.39
21	John Walker	.46	-1.01	.01	.75	-.16
22	Art Reynolds	3.97	-.77	-.13	.04	-.27
23	Jim Anderson	.47	.01	1.43	.35	-.30
24	Brian Patz	.39	.35	-.81	-.82	-1.24
25	Peter Carling	-.14	-1.32	.18	.15	.08
26	Bill Dobson	.04	-.10	-.41	.08	.23
	Eigenvalues	27.45	17.08	14.35	8.09	5.54
	% Total Variance	49.51	19.20	13.56	4.31	2.02

FIGURE 1


EIGENVALUE
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
-4
-6
-8
-10
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
DIMENSION

dimensions following a separate multidimensional scaling for each half. In the present study the total sample of 46 subjects was subjected to three such split half analyses. The solutions for all six subsamples were rotated to an orthogonal procrustean criterian using the total sample matix as a criterion. The first analysis split subjects into even vs. odd numbered halves while the second assigned subjects to two halves on the basis of a random number table. The third split half analysis was not random; the graduate student subjects (n=27) became one subsample and the undergraduate ones (n=19) another.

The correlations among the first five dimensions in the odd-even comparison were: .84, .98, .91, .94, and .75 respectively. A canonical correlation, using the dimensions of the even numbered subsample as one set of (predictor) variables and the dimensions of the odd numbered subsample as the other set of (criterion) variables, was performed to assess the similarity of the two subsamples as a whole. The resulting coefficient was .999 ($\underline{X}^2$ = 9999, df = 100, $p < .0001$). While the canonical correlation is a liberal test of the similarity of the two subsample configurations since it generates maximal variable weightings, the extremely high (canonical) similarity together with the high zero order correlations of corresponding dimensions attests to the stability of the dimensions in the odd-even subsample comparison.

The correlations among the first five dimensions in the randomly assigned subsamples were: .63, .97, .73, .79, and .69 respectively. The canonical correlation between these two random subsample matrices was .998 ($\underline{X}^2$ = 9999, df = 100, $p < .0001$). Again the split half analysis shows the very substantial similarity of the

configurations of the two subsamples even if the canonical correlation is taken to be somewhat of an overestimate.

The correlations for the first five dimensions between the graduate student subsample and the undergraduate subsample were: .63, .96, .74, .83, and .71 respectively. The canonical correlations between these two planned subsamples was .997 ($\underline{X}^2$ = 9999, df = 100, $p < .0001$). The high similarity in this comparison argues even further for the stability and generalizability of the dimensions given the subject differences in these two subsamples.

Not surprisingly, comparisons between the original (N=46) dimensions and the six split half subsamples were also extremely high. All canonical correlations exceeded .99. All zero order correlations between respective dimensions of the original configuration vs. the subsample configuration exceeded .61. The mean of all such correlations was .873. In short, there is considerable evidence for the stability of the original dimensions. Thus, we may now turn to their interpretation.

Interpretation of the Dimensions

The interpretation of the five rotated dimensions in the main configuration was made by examining the stimuli which projected .75 or higher at both poles of the dimensions. These salient stimuli were compared for their similarity. A summary of the interpretation is presented in Table 2.

Dimension I (Mental Illness)

The first dimension is clearly defined by Art Reynolds, the

Table 2

Summary Interpretation of the Rotated Dimensions

I	II	III	IV	V
		LABEL		
Mental Illness	Social vs. Psychological Deviance	Depression vs. Impulse Control	Anxiety vs. Impulse Control	Paranoid vs. Impulse Control
		Salient Positive Projections		
Insane (person)	Alcoholic	Depressive Neurosis	Nervous (person)	Schizophrenia, paranoid type
Schizophrenia, simple type	Ex-convict	Manic-Depressive illness, depressed type	Anxiety Neurosis	Schizophrenia, paranoid type not presently psychotic
	Drug addict	Manic-Depressive illness, depressed type, mild	John Walker	Alan Hill
	Alcoholism episodic excessive drinking	Jim Anderson		
	Jack Cole			
		Salient Negative Projections		
Art Reynolds	Schizophrenia, simple type in remission	Ex-convict	Ex-convict	Brian Patz
	John Walker	Jack Cole	Jack Cole	Drug addict
	Peter Carling	Brian Patz	Brian Patz	(Alcoholic)
				(Alcoholism episodic excessive drinking)

"normal" target at one end; the label "insane(person)" and the diagnostic category "schizophrenia, simple type" were the anchor points at the other end of dimension one. This dimension might therefore be termed a general "Psychopathology" dimension. Given the predominance of the "insane person" projection perhaps this dimension may be better termed "mental illness" reflecting the somewhat pejorative lay conception which may be related to social desirability, rather than a term (i.e., "Psychopathology") which has a more scientific connotation.

This interpretation is supported by the smaller projections on the dimensions. "Former psychiatric patient" and "psychiatric hospital patient," terms which denote some social stigma, also have notable projections at the appropriate end of the dimension. Furthermore, stimuli which are only indirectly related to psychopathology and which are thus probably less socially undesirable (i.e., alcoholic, ex-convict, nervous person, and "alcoholism, episodic excessive drinking"), fell on the normal side of the dimension. Similarly, the target descriptions defined by health concern, reticence and being unemotional, and mild depression (i.e., John Walker and Jim Anderson) were closest to Art Reynolds on this dimension.

Recall that this first dimension explains significantly more of the variance than any other dimension and about half of the total variance in the space. This supports the major hypothesis that judgments of psychopathology can be largely conceptualized as a generalized, mental disorder severity dimension related to social undesirability. This has therefore been labeled the "Mental Illness" dimension.

Looking at this first Mental Illness dimension with regard to the secondary hypotheses, it can be first noted that in general the stimuli with the most (specific) information were perceived as the most normal. The mean projection for the stimuli which were categorized as lay labels was -.36 while the mean for the diagnostic categories, which provided somewhat more specific information, was -.21. The brief personality descriptions, which provided the most information, characterized by specific behavioral exemplars, had a mean projection of +.64. Even when the salient anchor points (6,9, and 22) for dimension one are removed this finding holds. The mean projection for the three categories of stimuli in this case are -.06, -.83 and +.14 respectively.

The prediction that qualifying phrases used by the DSM II would not influence judgments was not clearly supported by the projections on dimension one, although a better case can be made for such a prediction with respect to the other dimensions. Thus, for example, the phrase "in remission" when paired with the stimulus "schizophrenia simple type" resulted in a considerably smaller projection on the first dimension. However, this effect was not repeated on the other dimensions. To a lesser degree the same pattern of effects occurred for the qualifiers "not presently psychotic" and "mild". Such qualifiers thus seem to influence judgments only with respect to our general perceptual categories, such as severity of disorder ("Mental Illness").

Finally, in support of the hypothesis of the ineffectiveness of qualifying psychopathology phrases it can be noted that "former psychiatric patient" and "psychiatric hospital patient" projected equally on dimension one. Apparently the addition of the word "former"

is not critical in this sample even though it may well be clinically important to distinguish former from current psychiatric status.

Dimension II (Social vs. Psychological Deviance)

Given the differential pattern of projections of the key stimuli described above, it is important to interpret the remaining dimensions despite their relatively smaller explanatory utility. The salient projections at the positive end of dimension two included: alcoholic, ex-convict, drug addict, "alcoholism, episodic excessive drinking" and the Jack Cole personality profile. Jack Cole may be best summarized as being psychopathic (rebellious, impulsive, antisocial etc.) and perhaps somewhat similar to the ex-convict stimulus. Taken as a whole this pole reflects a social deviance pattern.

The other pole of dimension two is anchored by the "schizophrenia simple type, in remission", John Walker and Peter Carling stimuli. John Walker was characterized by his health concern, unemotionality and lack of insight; Peter Carling was a similar person though he was not lacking in insight but instead suffered from "perceptual distortions". This pole is harder to conceptualize but reflects a mild psychopathology or schizophrenia pattern. Perhaps dimension two can best be summarized as a Social vs Psychological Deviance dimension.

Dimension III (Depression vs. Impulse Control)

Dimension three can be readily labeled a Depression and/or Impulse Control dimension. All three diagnostic categories related to depression, irrespective of qualifiers, projected very highly at the

positive pole. "Depressive neurosis" was not differentiated from the other related stimuli. Furthermore, the description characterized by the person who was feeling "very downhearted and 'blue'...," i.e., Jim Anderson, also defined this pole.

At the opposite end were stimuli characterized primarily by impulsivity and to some extent an antisocial quality. Thus, ex-convict and Jack Cole are again paired on a dimension. In addition the Brian Patz stimulus projected on this pole of dimension three.

Dimension IV (Anxiety vs. Impulse Control)

Dimension four is an Anxiety vs. Impulsivity factor. The positive pole is anchored by the "nervous person," and "anxiety neurosis" stimuli as well as the John Walker description which included a reference to "becoming anxious." The salient projections of the negative pole are identical to those just described for dimension three.

Dimension V (Paranoia vs. Impulse Control)

The positive pole of dimension five is clearly related to Paranoia or Ideas of Persecution. Thus, both relevant diagnostic categories and the Alan Hill description, which is characterized by cynicism and ideas of persecution, project strongly on this pole. Note that the qualifying phrase "not presently psychotic" changed the magnitude but not the salience and direction of the "schizophrenia, paranoid type" stimulus.

The negative pole is more difficult to interpret. It is anchored by the Brian Patz stimulus, reflecting the Impulsivity and

Neurotic Disorganization scales of the DPI. Even inclusion of less salient projections such as "drug addict" and the two alcoholic stimuli do not greatly enhance interpretation of this pole. The common factor of this pole may simply be its lower level of socially undesirable deviance, relative to paranoia, and so, Dimension five is tenuously labeled Paranoia. The reader may note that it may also be interpreted in a way similar to the Social-Psychological Deviance dimension, although this is perhaps even more tenuous than the Paranoia interpretation.

Part II: Analysis of the Ratings

Interpretation of the dimensions may be facilitated by an examination of the unidimensional stimulus ratings. The principle factor analysis of the stimulus by rating question matrix is depicted in Table 3. This table presents (procrustean) rotated loadings of the stimuli on factors derived from the rating questions scores of the subjects rather than their similarity judgments. The orthogonal procrustean rotation used the multidimensional scaling configuration as a hypothesis matrix.

The rotated factor structure appears very similar to the multidimensional scaling solution. Comparisons of the first five factors and dimensions yielded correlations of .654, .859, .757, .662, and .684 respectively. The same comparisons using Harman's (1976) coeficient of congruence produced virtually identical results: .633, .859, .756, .653, and .683 respectively. The canonical correlation between these two configurations was .998 (x^2 = 186.7, df = 90, $p < .0001$). Given the great convergence in the underlying structure of

Table 3

The Procrustean Rotated Factor Matrix of Psychopathology Descriptors Based on the Unidimensional Ratings

	Stimulus	Factor				
		I	II	III	IV	V
1	Former psychiatric patient	-.13	.12	-.06	-.46	-.09
2	Alcoholic	-.17	.26	.71	.31	-.39
3	Ex-convict	.13	.91	-.16	-.02	-.06
4	Psychiatric hospital patient	-.55	-.50	.29	.07	.49
5	Drug addict	-.59	.34	.55	.09	-.25
6	Insane (person)	-.94	-.26	-.08	-.10	-.02
7	Mentally disturbed (person)	-.47	-.63	.31	-.05	.17
8	Nervous (person)	.84	-.09	.16	.25	-.39
9	Schizophrenia, simple type	-.79	-.42	-.15	-.18	.10
10	Schizophrenia, simple type, in remission	-.45	.15	-.54	-.45	-.39
11	Schizophrenia, paranoid type	-.60	.00	-.40	.17	.50
12	Schizophreina, paranoid type not presently psychotic	-.24	.50	-.80	-.15	-.14
13	Anxiety neurosis	.64	.22	.21	.51	-.22
14	Depressive neurosis	.31	-.76	.29	.13	.29
15	Manic-depressive illness, depressed type	-.53	-.63	.12	-.19	.37
16	Manic-depressive illness, depressed type, mild	-.48	-.60	.04	-.57	-.13
17	Marital maladjustment	.77	.18	.28	.04	-.40
18	Alcoholism, episodic excessive drinking	.27	.27	.70	.28	-.42
19	Alan Hill	-.57	.20	-.54	-.07	.41
20	Jack Cole	.14	.90	-.16	.02	.33
21	John Walker	.08	-.48	-.20	-.21	-.25
22	Art Reynolds	.86	.03	-.19	.16	.12
23	Jim Anderson	.12	-.81	.35	-.07	.39
24	Brian Patz	.26	.55	-.36	-.04	-.24
25	Peter Carling	-.84	-.23	-.05	-.20	.05
26	Bill Dobson	-.45	.32	-.13	-.21	-.06
	Eigenvalues	10.07	6.38	3.26	2.33	2.00
	% Total Variance	38.7	24.5	12.5	8.9	7.7

Table 4

The Varimax Rotated Factor Matrix for the Rating Questions

	Rating	Factor				
		I	II	III	IV	V
1	Confidence	.05	-.16	.96	.21	-.03
2	Psychopathology	-.93	.16	-.05	-.31	.01
3	Institutionalization	-.92	.23	.01	-.25	.13
4	Base Rate	.75	-.15	.49	.09	-.37
5	Social Undesirability	-.74	.05	-.09	-.31	.55
6	Passive/Active	.25	-.91	.10	-.21	.15
7	Stable/Changeable	-.41	.04	-.17	-.87	-.01
8	Good/Bad	-.45	.04	-.26	-.71	.37
9	Sane/Insane	-.94	.13	-.09	-.26	.01
10	Weak/Strong	.10	-.88	.14	.35	-.23
	Eigenvalues (Unrotated)	5.89	1.52	1.18	.68	.35
	% Total Variance	58.90	15.23	11.78	6.76	3.53
	Rotated Eigenvalues	4.12	1.75	1.32	1.80	.67
	% Common Variance	42.65	18.12	13.66	18.63	6.94

the stimuli based on multidimensional analysis of the similarity judgments (Table 1) as compared to the factor analysis of the unidimensional ratings (Table 3), it is useful to interpret this structure in terms of the factors underlying the rating questions themselves.

Table 4 depicts the rotated, principle components factor matrix for the analysis of those ratings. Only the first four factors were interpretable and only the first one has any great significance for purposes of interpretation. The positive pole of the first factor reflects judged frequency of occurrence or base rate. The opposite pole clearly appears to represent judgments of pathological severity and social desirability and thus we can label this factor "Severity - Prevelance." Since this is the largest factor of the rating questions we can say that the subjects' ratings of the stimuli appear to be largely accounted for by a Severity and Prevelance factor which is reminiscent of the large "Mental Illness" dimension described earlier. Factor two may be a semantic factor since it is defined by the passive-active and weak-strong semantic differential items. Factor three appears to be essentially a rating confidence factor, given the high loading of the first rating question. The stable-unstable and good-bad ratings appear to be the only major variables loading on factor four.

DISCUSSION

The five multidimensional scaling dimensions which were interpreted accounted for more than 95% of the total variance. Most of

the variability in the judges' conceptualizations of the psychopathology descriptors, in other words, can be explained by these dimensions. Furthermore, the stability of these dimensions over different samples permits us to interpret these dimensions as meaningful and generalizable representations of the major conceptualizations used in clinical judgments of psychopathology.

As hypothesized the primary dimension underlying the conceptualizations of the judges was a large, general, dimension, which was labeled "Mental Illness" because it reflected both psychopathological severity and social desirability. It is the pejorative, stigmatizing elements of this dimension (e.g., insane person) which make the label "Mental Illness" more appropriate than a term such as "psychopathology."

While this first dimension explains more than half the judgmental variance, the second dimension was also quite important; it accounted for about a fifth of the variance and was labeled "Social vs. Psychological Deviance." "Impulse Control" was a major component of each of the remaining three dimensions, a finding which is consistent with results reported by Chan and Jackson (1978). The other components of these "Impulse Control" dimensions were Depression, Anxiety and Paranoia, respectively.

The factor analyses of the unidimensional stimulus ratings further substantiate the interpretation of these dimensions. This is most succintly demonstrated by the good correspondence between the dimensions and the unidimensional stimulus factors. Method variance appears to make the biggest difference on the Mental Illness dimension. The factor analysis of the rating questions also produced a large, first

factor which was called "Severity-Prevelance," a term which bears great similarity to the Mental Illness dimension. In this sense there is some convergent validity for this Mental Illness interpretation inherent in these findings.

As expected, it appears that labels and diagnostic categories function very similarily with respect to this first dimension. Thus, it does not appear that the taxonomy of the DSM II, while on the face more specific than labels, is perceived so by judges.

This has important implications for clinical practice since there was evidence to support the hypothesis that more specific stimuli were perceived as being relatively normal. If informational specificity of psychopathology descriptors is inversely related to broad, evaluatively based judgmental dimensions, then a strong case can be made for behaviorally oriented approaches to psychopathological nosology. One such approach to behavioral classification, i.e., the modal profile descriptions, was exemplified in this study.

Finally, with respect to the labeling issue there was also some evidence that qualifying phrases such as those used in the DSM II had minimal impact on clinical judgments. In other words the presence of moderating qualifyers (e.g., "mild," "in remission") are not very effective in overcoming the powerful biasing effects manifested by labels and diagnostic categories, even though this may have been their purpose. This issue is relevant to the recent controversy about the sensitivity of psychodiagnostic judgments to subtle patient differences. There is debate for example, about the importance and implications of the phrase "in remission," (Farber, 1975; Rosenhan, 1975; Spitzer, 1975).

In the present study qualifying phrases had a mild, moderating

effect for stimuli on the first dimension, which is encouraging. However, this effect did not occur on any of the other dimensions. Given this and the minimal effectiveness of such qualifiers on the major clinical judgment dimension one must conclude that qualifying phrases may not at all function, diagnostically, as denotative (or connotative) moderators of a patient's clinical status.

In conclusion, the present study does not claim to have investigated an exhaustive set of psychopathology descriptors. Nonetheless, the stimuli chosen were sufficiently representative and the obtained dimensions sufficiently stable and interpretable to serve as an important step in mapping the dimensions of psychopathology inherent in clinical judgments. Future research might aim to replicate the large "Mental Illness" type dimension, discovered in this study, with different judges and with expanded sets of stimuli. The importance of replicating this dimension is supported in this study by the finding that the multidimensional scaling and factor analytic approaches, while quite similar in their results, were most discrepant in the case of this large Mental Illness dimension. While the canonical correlation analysis used to compare these two approaches revealed an extreme similarity for the two configurations as a whole, the simple correlations between corresponding dimensions and factors were not so high. A different set of rating questions may produce a higher correspondence. The stability of the five dimensions observed in this study suggests that future research can adopt these as an adequate launching point to the specification of the important dimensions that judges use to evaluate psychopathology.

CHAPTER THREE

The Influence of diagnostic suggestion and source credibility on labeling effects in inferential accuracy

The purpose of this experiment is to examine the effects of diagnostic suggestion and source credibility, two variables often implicated in clinical judgment, persuasion or attitude change, and person perception studies (e.g., Adinolfi, 1971; Di Nardo, 1975; Gordon, 1976; Lee & Temerlin, 1970; Scheid, 1976; Temerlin, 1968, 1970; Temerlin & Trousdale, 1969; Yaffe, 1972). This study examines the effects of these two variables on the parameters of the model of Inferential Accuracy (Jackson, 1972; Reed & Jackson, 1975). In view of the apparent potency of these variables it is expected that important interactions will be discovered when differing labels and patterns of psychopathology are involved in a judgment task which manipulates levels of diagnostic suggestion and source credibility. The complexity of these variables is illustrated by studies which demonstrate that the presence of labels, information about social class, demographic characteristics or previous history, diagnostic suggestions about pathology, and the credibility status of the attributor (i.e., "source credibility") all have a direct and perhaps an interactive effect in biasing clinical judgments.

With respect to diagnostic suggestion and source credibility, it has been shown that the mere suggestion of the presence of psychopathology by high prestige sources such as psychiatrists and psychologists is sufficient to induce powerful biases in clinical judgments even in the face of objective evidence to the contrary (Di Nardo, 1975; Lee &

Temerlin, 1970). Most studies employ ratings of patients as dependent measures (Abramowitz & Dokechi, 1977) and do not directly assess the impact of such biases on parameters of clinical judgment such as accuracy of diagnoses or predictions.

The present study on the other hand, independently manipulates labels, diagnostic suggestion, and source prestige or credibility and evaluates their effects on clinical judgments in terms of the model of inferential accuracy as well as on patient rating measures. The design avoids the problems in earlier studies (e.g., Di Nardo, 1975; Lee & Temerlin, 1970; Temerlin, 1968,1970; Temerlin & Trousdale, 1969) of the confounding of diagnostic suggestion with valid information, the credibility of the source of the diagnosis and other important variables in clinical judgments.

METHOD

Design Overview

The experimental paradigm and the subjects' task is similar to that described by Paquin and Jackson (1977). Each subject or "judge" evaluates four target descriptions which may have labeling information associated with them. The major manipulation however, results in the subject receiving either a high or a low suggestion about the presence of debilitating psychopathology as well as information reflecting one of three levels of source credibility, here operationalized as professional status or prestige. The subject is then asked to predict the DPI responses of the described person and again the subject's performance is evaluated in terms of the threshold and

sensitivity dependent variables (Inferential Accuracy) as well as prognosis judgments largely identical to those involved in the preceeding study.

Threshold is operationalized as the judge's mean response across the 52 DPI items, for each target; Sensitivity is the correlation between the (two item) DPI scale scores predicted by the judge and the standardized loadings of the modal profile over the DPI scales. Thus, sensitivity represents the judge's ability to predict the responses of actual patient types. Sensitivity scores will be underestimates of accuracy in this study because of the small number of items (2) determining DPI scale scores.

Thus, the design can be summarized as a 4 X 2 X 3 X 2 X 2 (targets X labels X source X suggestion X sex) fixed factor, repeated measures (targets) analysis of variance design, with subjects nested within all the variables except targets (i.e., the description). The lack of exposure to the varying levels of the other variables helps to minimize the demand characteristics of the experiment.

Subjects

Volunteers (45 males and 74 females) receiving course credit in undergraduate psychology classes for their participation comprised the subject sample for this experiment.

Procedure

Subjects were informed that the purpose of the study was to

"examine processes in making decisions about clinical judgments and that we can improve the accuracy of such judgments by enhancing our culturally learned, largely accurate, conceptions about the relationships among various behavior patterns and traits." The actual instructions and answer sheets used by the subjects are outlined in Appendix B.

Subjects were randomly assigned to either the labeled or unlabeled condition. In the latter case the description read "normal medical background" under "History." The target descriptions in the labeled condition contained information identifying their history as "former psychiatric patient," and provided a few basic details about this previous status. Aside from this, each of the descriptions contained one combination of the two levels of diagnostic suggestion by three levels of source credibility status. In order to minimize any uncontrolled "set" for psychopathology the descriptions were presented as brief summaries of a "routine psychological assessment as part of a large company's employee benefits program." Representative samples of the Assessment Reports, containing the target descriptions, and the labeling, suggestion and source manipulations, are presented in Appendix B, which also contains the DPI item information.

Diagnostic Suggestion

The target descriptions contained indications of either the high or low probability of the presence of serious debilitating psychopathology. This information was presented in the concluding statement

of the target description. Since the modal profile types, from which the descriptions are derived, are defined more by shape than elevation these suggestions can reasonably take the form of statements about symptom severity. A conclusion about overall diagnosis with respect to "psychosis" is not given thus departing from the format of previous studies (Di Nardo, 1975; Lee & Temerlin, 1970; Temerlin, 1968). It is possible that gross terms like "psychosis" or "neurosis" impart trait information (e.g., hallucinations, bizarre thinking) which, despite their vagueness, may confound a simple, severity suggestion. The exact wording of the diagnostic suggestion varies slightly for each target in order to maintain the credibility of the source.

Thus, a typical high diagnostic suggestion would read:

> "In conclusion, the above pattern of traits reflects a high likelihood of serious psychopathology and is probably resulting in severely debilitating consequences for this person. While a singular diagnosis is difficult the above pattern is interesting in that this person may often appear normal but probably suffers from these symptoms frequently."

A typical, low diagnostic suggestion would read:

> "In conclusion, the above profile of characteristics for this person represents only a minimal likelihood of psychopathology severe enough to be incapacitating. While some uncertainty exists, as always, this person can be diagnosed as only rarely experiencing these difficulties."

Source Credibility Status

Unlike previous studies, the diagnostic suggestion associated with a clinical judgment is manipulated independently of the degree of credibility or status (prestige) of the person making the suggestion.

Judges did receive target descriptions identified with either high ("Psychiatrist (MD)"), medium ("Social worker (MSW)"), or low ("Secretary-clerk") source credibility status. The status of these occupational roles was previously ascertained in a study of 21 senior undergraduates, which asked the subjects to rank order a list of 9 such roles in terms of "their competence to make important diagnostic decisions and clinical judgments about patients." The results of this ranking is depicted in Table 1. The source credibility manipulation was further verified by cross validating these rankings in an independent sample of 10 introductory psychology students. The correlation between the mean, rank order, ratings for these two groups was .96 ($p < .001$).

Experimental Stimuli

Three of the targets depicted empirically identified psychopathological types developed by an approach termed "modal profile analysis" as described in Skinner, Jackson, and Hoffman (1974) and Skinner, Reed, and Jackson (1976). The target descriptions reflect the salient DPI scale scores of the profile with the addition of the label "former psychiatric patient" in the labeling condition. These three targets were selected according to their base rate of occurrence averaged over nine separate clinical and college populations (five hospital, three college and a prison sample) totalling 953 tested individuals (Skinner et al., 1976).

Thus, a high, medium and low base rate target was selected according to the overall prevalence of the profile (target) in the total

Table 1

Rank Order of Selected Mental Health Professionals

Occupation	Mean Ranking[a]	Obtained Rank[b]
General Practitioner (MD)	4.81 (1.85)	4
Nurse, in psychiatry (RN)	5.76 (1.49)	7
Psychiatrist (MD)	1.33 (.26)	1
Psychiatry Intern	3.67 (1.36)	3
Psychologist (PhD)	1.81 (.59)	2
Psychometrist (BA)	5.76 (1.38)	6
Secretary-Clerk (in psychology department)	8.67 (.64)	9
Social Worker, in psychiatry (MSW)	4.90 (1.23)	5
Undergraduate student (doing a testing course project)	7.14 (1.86)	8

[a] Based on a sample of undergraduates enrolled in an abnormal psychology course (n = 21). The numbers in parentheses are standard deviations.

[b] The lower the rank the more credibility is associated with the occupational source. Thus, the psychiatrist was judged the most credible (competent).

Table 2

Characteristics of the Four Target Descriptions

Name	Modal Profile[a]	Defining DPI Scales	Prevalence[b]	Mean Ranking[c]
1. John Walker	1+	Health Concern, Repression, Shallow Affect	239 (25.3%)	1.0
2. Jim Anderson	3-	Desocialization Repression, Self Depreciation, Depression	87 (9.2%)	2.6
3. Alan Hill	5+	Cynicism, Ideas of Persecution, Sadism, Shallow Affect	16 (1.7%)	5.6
4. Art Reynolds	Normal	None	-------	-------

[a] This column refers to the empirically identified, profile type which defined the targets; they are derived from a multiprofile analysis of five clinical and three college populations described in Skinner, Reed, and Jackson (1976). The profile types used in this study were developed in a study by Reed (1976).

[b] The number of people classified by the given profile type in the multi sample (N = 953) population of Skinner, Reed, and Jackson (1976).

[c] The average rank order of the profile across eight samples (Skinner, Reed & Jackson, 1976).

sample. This is shown in Table 2: the mean ranking across the nine samples is also shown in the table. These figures can be taken as an estimate of the prevalence of the pattern of psychopathology, reflected by the profile type, in the general population and hence the degree to which subjects will be familiar with them. The fourth target is intended as a relatively normal person, lacking in obvious psychopathology as described in Reed and Jackson (1975).

The subject judged the four targets in terms of 52 items selected from the DPI (see Appendix B) according to criteria detailed in Paquin and Jackson (1972, p.111). Finally, each judge rates the targets according to several clinical or prognostic judgments on nine point scales, as well as a number of semantic differential ratings also on nine point scales. The clinical judgment questions involve estimates of degree of psychopathology, need for institutionalized care, therapy prognosis, judged frequency of occurrence, social desirability etc., as detailed in Appendix B.

In summary, this study extends the examination of labeling effects on clinical judgments by manipulating diagnostic suggestion and source credibility. It is hypothesized that all these variables would lower inferential accuracy but that such an effect would be greatest in the high diagnostic suggestion, high source credibility conditions. Furthermore, target descriptions will produce interactions with these independent variables since the targets reflect differential base rates and therefore the judges' degree of familiarity with the underlying patterns of psychopathology which are represented by the targets.

RESULTS

Correlations among the 19 dependent variables revealed a moderate degree of relationship. Most of the correlations in the matrix fell in the $\pm$.40 range although a few were as high as .76. To clarify the underlying pattern of relationships of the dependent variables this matrix was factored by principle components and subjected to orthogonal procrustean rotation. Seven components, accounting for 71% of the total variance, were extracted and rotated; these are reported in Table 3.

The ability of this greatly reduced number of factors to account for most of the variance suggests that the 19 dependent variables can better be analyzed in terms of the reduced rank solution. This was accomplished by obtaining a set of orthogonal component scores based on the procrustean rotated factor matrix.

These component scores were obtained by matrix operations which transformed the original variables into orthogonal, composite variables based upon the principle components indicated in Table 3.

$$Y = ZA(A'A)^{-1}$$

Specifically, where Y is the principle component (or factor) score, Z is the standardized (M = 0, s.d. = 1) subjects by variables matrix of original, dependent variables, and A is the rotated, principle component matrix (Horst, 1965). The rotation of this matrix is an orthogonal procrustean one which used as the hypothesis matrix the salient loadings indicated in Table 3. The hypothesis matrix was derived from a varimax rotated solution obtained earlier.

Table 3

Rotated Principle Components Solution for the Dependent Variables

	I	II	III	IV	V	VI	VII
Threshold	.027	.101	.494*	-.019	-.055	.515	.238
Sensitivity	.038	.063	-.197	.892*	.098	-.019	.146
Confidence	.140	-.341	.038	.000	.521	.424	-.608
Psychopathology	.760*	.105	.363	.063	.119	.199	.162
Institutional Care	.762*	.110	.271	-.003	.089	.012	.167
Base rate/1,000	-.250	-.005	-.510*	.051	-.060	-.027	-.426
Social Undesirability	.692*	-.054	-.164	-.197	-.051	.079	.124
Therapeutic Benefit	.291	.179	.743	.068	.130	-.203	-.192
Recidivism	.472*	.074	.067	.095	-.238	.541*	.209
Passive-Active	-.146	.147	-.738*	.171	.210	-.087	-.203
Stable-Changeable	.350	.456*	.228	-.048	.049	.434*	.206
Good-Bad	.427	.104	.133	.072	.373	.172	.508*
Sane-Insane	.670*	.167	.327	.080	.168	.167	.311
Weak-Strong	-.149	-.084	-.722*	.024	.226	-.234	-.183
Mature-Youthful	.176	.167	.365	-.066	.133	.435*	.466*
Sensitive-Insensitive	.067	-.235	-.064	.304	.626*	.024	.474
Excitable-Calm	-.052	-.800*	.006	-.112	-.142	-.263	.110
Cautious-Rash	.020	.427	-.237	-.206	.644*	-.079	.248
Warm-Cold	.369	-.174	.360	.139	.216	.275	.533*
% Total Variance	31.69	11.33	7.01	6.23	5.21	4.87	4.40
% Common Variance	44.80	16.01	9.91	8.81	7.36	6.88	6.22

* Salient variables used in defining the hypothesis matrix for the procrustean rotation.

Several analyses were employed to ensure the stability of these new component scores. These involved rotating the principle components to alternative, theoretically meaningful hypothesis matrices and comparing the similarity of the resulting factor structures. The canonical correlations between the original, varimax rotated, components and these alternative solutions exceeded .98 in all cases (reported below) and these were significant beyond the .0001 level.

The first of these comparisons decomposed the raw variables by a factor analytic rather than the principle components solution, reported in Table 3. These factors were then used as a hypothesis matrix to rotate the principle components and this rotated matrix was compared to the original components matrix. Harman (1976) describes a coefficient of congruence measure for making such comparisons. This coefficient, which is designed specifically to assess factor matrix similarity, has a range of -1.00 to +1.00 and is distributed very much like the correlation coefficient although it is more conservative. The coefficients of congruence for the respective factors of this first comparison were: .999, .997, .990, .982, .959, .959, .887, .723, .686, and .615 over the first ten extracted factors.

Using the same approach the principle components were next rotated to a hypothesis matrix based upon the factoring of the dependent variables for the 189 subjects in the study reported in Chapter four. The coefficients of congruence for the first ten factors were: .947, .948, .923, .947, .954, .838, .754, .751, .923, and .861 respectively. The stability of the factors is somewhat reduced after seven factors in a number of such comparisons. This fact together with the small percentage

of variance accounted for by the residual factors led to the utilization of only the first seven original principle components.

In summary there are seven new, uncorrelated, dependent variables of considerable stability produced by this method, which are a linear combination of the original variables. These new variables were interpreted (below) and subjected to analyses of variance. There are several advantages to this approach. Unlike multivariate analysis of variance, which yields an overall F ratio indicative of significant effects somewhere in the set of dependent variables, it provides a meaningful interpretation of the effects directly. Secondly, the composite variables provide more reliable dependent variable scores. Finally, the reduced set of mutually orthogonal scores minimizes the problem of inflated alpha levels. The results of the analyses of variance using this approach are described below. Normality assumptions were tested by examining the sampling error of measures of kurtosis and skewness for all variables within each target. Virtually all such tests revealed that the variables involved were normally distributed. In the few cases where this was not the case stricter alpha levels were adopted. Similarily, homogeneity of variance tests revealed that virtually all estimates of population variances within targets were equal. Stricter significance tests were adopted for the few cases, involving the third target, where heterogeneity of variance was suspected.

Component 1 Mental Illness:	+.76 Psychopathology	+.67 Sane-insane
	+.76 Institutional care	+.47 Recidivism
	+.69 Social Undersirability	+.43 Good-bad

The first dependent variable is termed "mental illness." In

addition to loadings for rated psychopathology and need for institutionalized care this variable is characterized by desirability determinants with a pejorative connotation: good-bad, sane-insane, recidivism, and judged social undesirability. Thus, the interpretation of this variable is very similar to that given for the first dimension and factor in the preceeding study. Table 3 shows that this mental illness factor accounts for about 32% of the total variance and about 45% of the variance common to the seven rotated components.

Table 4 reveals that the analysis of variance (ANOVA) for this mental illness variable yields significant effects for diagnostic suggestion, targets and many of the interactions. However, it seems that these effects vary greatly in the amount of variance they explain. Using $\hat{\omega}^2$ (omega squared) as an estimate of the percentage of variance explained by an effect, targets and suggestion accounted for about 14% and 17% of the variance respectively. Suggestion explains about half the variance, according to an alternative estimate. The remaining main and interaction effects accounted for considerably less of the variance.

A measure of variance accounted for, namely, $\hat{\omega}^2$ was calculated according to formulae outlined in Keppel (1973, pp. 552-553) and Vaughan and Corbalis (1973, pp. 208-210). The importance of $\hat{\omega}^2$ in terms of estimating the magnitude of treatment effects lies in the fact that the significance of an F ratio is determined by sample size so that even small effects can be highly significant if the sample is large enough. This is not the case with $\hat{\omega}^2$.

Unfortunately, $\hat{\omega}^2$ is an underestimate of the proportion of total variance for an effect in a non-additive repeated measures design,

Table 4

ANOVA Summary for Mental Illness

Effect	df	MS	F	$\hat{\omega}^2$[a]	$\hat{\omega}^2$MS[b]
Z	1,95	83.86	84.158**	.166	.510
GY	2,95	3.08	3.088	.010	.013
LY	2,95	2.92	2.933	.010	.012
GLZ	1,95	8.65	8.683*	.015	.047
YZ	2,95	6.24	6.261*	.023	.032
GYZ	2,95	5.01	5.024*	.018	.025
GLYZ	2,95	7.09	7.115*	.026	.037
S	95	.99			
T	3,285	23.80	57.305**	.142	.144
TY	6,285	.94	2.260	.010	.003
TZ	3,285	3.58	8.622**	.021	.019
GTL	3,285	1.52	3.648	.008	.007
TLY	6,285	.97	2.340	.011	.003
TYZ	6,285	1.16	2.782	.013	.005
GTLY	6,285	1.50	3.615*	.017	.007
GTLYZ	6,285	1.26	3.041*	.014	.005
ST	285	.41			

Index: G - Gender (Sex)
T - Target
L - Label
Y - Source
Z - Suggestion
S - Subjects

* $p < .01$.
** $p < .001$.

NOTE: All non-significant ($p > .05$) effects have been omitted.

[a] Omega squared, the proportion of explained variance

[b] $\{MS_i - MS_e\} \div \Sigma MS$ where i = effect of concern, e = error term

particularly when the repeated measure is a fixed factor (Vaughan & Corbalis, 1969). However, the ratio of the MS (mean square) for an effect minus MS error, divided by the sum of all MSs, i.e., $\{MS_i - MS_e\} \div \Sigma MS$ can be considered another, though positively biased, estimate of the relative magnitude of a variance estimate in a given experiment. This approach assumes, unlike Endler (1966), that the MS is an independent estimate of variance, given a true null hypothesis, composed of treatment effect plus error; highest order interaction effects are not assumed to be residual (error) components i.e., interaction $\sigma^2_{ijk} \neq 0$. Both $\hat{\omega}^2$ and $\{MS_i - MS_e\} \div \Sigma MS$ ($\hat{\omega}^2 MS$) values are therefore reported in Table 4. $\hat{\omega}^2$ MS tends to produce higher coefficients, particularly for "between subjects," effects than does $\hat{\omega}^2$ in this experiment.

Looking first then at the effects involving targets (see Table 4) a number of important findings emerge. First of all, as expected there was a powerful main effect for the targets, $F(3,285) = 57.31$, $p < .00001$, which accounted for 14% of the variance on this mental illness variable. The means for the four targets were: Alan Hill, the least prevalent target, .575; John Walker, the most prevalent target, -.182; Jim Anderson, .267 and Art Reynolds, the normal target, -.527. The target means will be reported in the same order from this point onwards.

Post hoc tests revealed that <u>all</u> paired comparisons of these means were significant ($p < .01$); all a posteriori comparisons follow the Duncan new multiple range procedure. This test is one of the most appropriate ones when both sensitivity in detecting real differences and minimization of Type I errors is desirable (Carmer & Swanson, 1973). The results of these pairwise comparisons among targets hold true even using Scheffe's more stringent a posteriori F test. Secondly,

there were significant interactions for targets with source ($p = .037$), and suggestion ($p = .00008$).

Table 4 also shows a major, main effect for diagnostic suggestion ($F(1,95) = 84.16$, $p < .00001$). Suggestions that the characteristics of the target were "serious" or "frequent" led to a higher mean mental illness score (.489) than in the opposite condition (-.423). This variable accounted for at least 16.6% of the variance depending upon the estimate used.

The target by source interaction revealed that mental illness judgments of the targets were differentially influenced by the source of the test report. The means for the within target pairwise comparisons revealed that John Walker and Jim Anderson were judged significantly ($p < .01$) less mentally ill in the psychiatrist condition while Art Reynolds was judged most mentally ill in that condition. The social worker source resulted in significantly higher means than the secretary-clerk for Jim Anderson and Art Reynolds.

Table 4 shows that the target by suggestion interaction was highly significant and explains about 2% of the mental illness variance. All within target comparisons revealed that significantly ($p < .01$) more mental illness was attributed to targets in the high suggestion condition, with the exception of Art Reynolds. The latter was, however, significant at the .05 level. The four target means in the high suggestion condition were: 1.09, 2.48, .934, and -.312 respectively, vs. .063, -.612, -.400, and -.741 in the low condition.

The source by suggestion interaction which accounted for more than 2% of the variance revealed that all pairwise compaisons were in the expected direction of higher means in the high suggestion condition

and were significant. The means in the high suggestion condition were: .373, .835, and .259 vs. -.492, -.530, and -.247 in the low suggestion condition for the psychiatrist, social worker and secretary sources respectively. The biggest difference ocurred for the social worker and the smallest for the secretary. In other words suggestion significantly increases mental illness scores, particularly when the credibility of the source is moderate.

Finally, a number of other interactions reported in Table 4 are significant. These seem to involve sex in many cases and this may explain the absence of main effects for sex, labels and suggestion. Note for example, the significant sex by source, and label by source interactions as well as the major interaction ($p = .00169$) of sex, label, source, and suggestion. The latter interaction with sex removed (i.e., label by source by suggestion) is not significant ($F < 1$).

Component II
Excitability-Impulsivity: -.80 Excitable-calm +.46 Stable-changeable +.43 Cautious-rash

The second dependent variable can be considered an impulsivity type factor based on the salient loadings of the bipolar ratings listed above. The excitable, changeable, and rash poles of the ratings load in a consistent fashion at the positive end of this component. Table 5 reveals that the only main effect which was significant is suggestion ($F(1,95) = 5.16$, $p = .024$). Thus diagnostic suggestion again produces a highly significant effect accounting for more than 1% of the variance in the "impulsivity" scores. The mean of the high suggestion condition was .105 compared to -.144 in the low condition. The significant ($F(6,$

Table 5: ANOVA Summary for Three Variables

Effect	df	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS
		Excitable/Impulsive				Prognosis/Threshold				Sensitivity			
L	1,95					14.92	21.66**	.033	.140				
Y	2,95					2.80	4.12	.011	.021				
Z	1,95	6.27	5.16	.011	.130								
GY	2,95									4.02	5.28*	.016	.041
GZ	1,95					3.63	5.35	.007	.029				
LZ	1,95					5.51	8.11*	.011	.048				
S	95	1.21				.67				.76			
T	3,285					46.68	100.89**	.320	.456	50.58	90.94**	.164	.636
GT	3,285					1.62	3.49	.010	.011				
TZ	3,285	2.39	2.61	.013	.038	3.34	7.22**	.001	.028	2.20	3.95*	.013	.021
GTY	6,285					1.17	2.53	.015	.007				
TLZ	3,285									2.82	5.07*	.017	.029
TYZ	6,285					.99	2.13	.013	.005				
GTLY	6,285					.98	2.11	.012	.005				
GTYZ	6,285					1.01	2.18	.013	.005				
TLYZ	6,285	2.80	3.05*	.033	.048					1.99	3.58*	.025	.018
ST	285					.46				.55			

* $p < .01$.

** $p < .001$.

NOTE: All non-significant ($p > .05$) effects have been omitted, $\hat{\omega}^2$MS is an alternative measure of explained variance ($\hat{\omega}^2$), based on $(MS_t - MS_e) + \Sigma MS$. G = Gender (sex of judge), L = Label, T = Target, Y = Source, Z = Suggestion, S = Subjects.

285) = 3.05, p = .0068) four way interaction of targets, labels, source and suggestion accounted for the most variance (4%) suggesting that excitability-impulsivity scores were complexly determined. Although targets were not significant (p = .079), the target by suggestion interaction approached significance ($F(3,285) = 2.61$, p = .0510). Within target pairwise comparisons revealed that significant differences ($p < .01$) occurred only for the Alan Hill and Art Reynolds targets; both received higher excitability scores in the high suggestion condition. The means for the four targets in the high diagnostic suggestion condition were: .131, -.218, .090, and .416. In the low suggestion condition the target means were: -.194, -.220, .075, and -.237 respectively.

Component III			
Prognosis-Threshold:	+.74 Therapeutic benefit	-.74 Passive-active	
	+.49 Threshold	-.72 Weak-strong	
		-.51 Base rate	

Rated likelihood of achieving benefit from therapy loads on the same pole of this dimension as threshold, the decision point for attributing psychopathology. Thus, high estimates of the likelihood of the target deriving benefit from therapy were associated with a relative willingness to attribute psychopathology. Supporting the notion of a prognosis threshold interpretation are the loadings at the opposite end; weakness, passivity, and patients suffering rarely occurring syndromes are not associated with therapeutic benefits in this conceptualization.

This prognosis-threshold variable was most influenced by the targets, which accounted for at least one third of the variance; the four means were: -.133, .457, .475, and -.972. The most dramatic comparisons are for the latter two targets (Anderson and Reynolds), who

manifested the best and worst judged prognosis respectively; all paired comparisons were significant ($p < .01$) except Jim Anderson (.487) vs. John Walker (.475) means.

Labeling effects were also significant beyond the .0001 level (Table 5) and account for 3 to 14% of the variance depending on the estimate. The means for the patient label vs. normal background histories were .149 vs. -.236 respectively. The main effect for source was also significant and accounts for less than 2% of the variance. The psychiatrist, social worker and secretary/clerk means were: -.041, -.189, and .100, respectively. All pairwise comparisons were significant ($p < .01$).

A number of important interactions are revealed in Table 5. The target by suggestion interaction accounted for some of this variance; diagnostic suggestions apparently have differential impact on targets. The means for the four targets in the high suggestion condition were: -.254, .379, .348, and -.719 respectively vs. -.012, 1.534, .601, and -1.225, in the low condition. All within target comparisons were significant ($p < .01$), however, the largest pairwise comparison is clearly Art Reynolds.

The label by suggestion interaction also accounted for a relatively large percent (>1%) of the prognosis-threshold variance. The patient label means were: .248, and .050 for the high vs. low suggestion condition, respectively, vs. -.371 and -.101, the normal background means. The high suggestion patient means were significantly ($p < .01$) different; the low suggestion means were not ($p > .05$). Furthermore, both the high suggestion means were significantly different ($p < .05$)

from all other means. Apparently, the patient label is influential on the prognosis scores only in the context of a high diagnostic suggestion.

Finally, in addition to the target by source by suggestion interaction ($F(6,285) = 2.13$, $p = .049$), there were three significant interactions, together accounting for more than 3% of the variance, involving sex: sex by suggestion, sex by target, as well as sex by target by source by suggestion, and sex by target by label by suggestion. In short, prognosis-threshold is complexly determined.

Component IV Sensitivity: +.89 Sensitivity score

Dependent variable four is defined as a factor reflecting the judge's ability to predict actual patient responses, i.e., sensitivity, a component of inferential accuracy. No other variables loaded on this component and so it is interpreted as a relatively pure sensitivity variable. It is interesting, however, that the sensitive-insensitive rating had a small loading (.30) on this factor. Apparently, judges are more sensitive, to a slight degree, to targets they rated as being insensitive.

The largest main effects on sensitivity, as before, are due to targets ($F(3,285) = 90.94$, $p < .00001$). The means for the targets were: .355, .465, -1.08, and .082. All pairwise comparisons among target means were significant ($p < .01$). Targets accounted for 16% of the sensitivity variance or, 64% of the variance according to $\hat{\omega}^2 MS$.

Within target comparisons of the target by suggestion

interaction revealed that only the Anderson and Reynold targets yielded significant differences ($p < .01$). However, diagnostic suggestion had the opposite effects on them; sensitivity scores for Art Reynolds were higher in the high suggestion condition but lower for Jim Anderson!

The complexity of the influences on sensitivity is illustrated by highly significant interactions such as the target by label by source by suggestion interaction which accounted for 2.5% of the variance in sensitivity scores. The target by label by suggestion interaction, which accounted for about 2% of the variance was also significant ($p = .0023$). A few of the significant within target pairwise comparisons for the latter are noteworthy. In the low suggestion, normal background condition Alan Hill obtained the lowest scores but Jim Anderson the highest. More importantly, judges were most sensitive to Art Reynolds, in the high suggestion, patient label condition.

Component V		
Impulsivity-Psychopathy:	+.64	Cautious-rash
	+.63	Sensitive-insensitive
	+.52	Confidence

The fifth component variable is defined by semantic differential ratings reflecting a lack of caution and sensitivity. The mild evaluative (good-bad) loading of .37 (see Table 3) suggests that this impulsivity may be similar in some ways to the evaluative nature of judged psychopathic characteristics. Judges tended to give higher

Table 6: ANOVA Summary for Three Variables

Effect	df	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS
		Impulsivity/Psychopathy				Threshold/Recidivism				Social Undesirability			
GL	1,95					13.53	7.67*	.024	.221				
GLZ	1,95									7.03	4.63	.012	.094
S	95	1.52				1.76				1.51			
T	3,285	23.73	39.49**	.151	.396	6.00	9.11**	.036	.100	29.96	53.22**	.200	.500
TY	6,285					1.87	2.84	.022	.023				
TZ	3,285									2.01	3.56	.012	.025
GTY	6,285	1.35	2.25	.016	.013	1.59	2.41	.018	.018				
TLYZ	6,285	1.82	3.03*	.022	.021								
ST	285	.60				.66				.56			

* $p < .01$.

** $p < .001$.

NOTE: All non-significant ($p > .05$) effects have been omitted. $\hat{\omega}^2$ MS is an alternative measure of explained variance ($\hat{\omega}^2$), based on $\{MS_i - MS_e\} \div \Sigma MS$. G = Gender (sex of judge), L = Label, T = Target, Y = Source, Z = Suggestion, S = Subjects.

confidence ratings to the targets they judged as being rash and insensitive.

The ANOVA summary in Table 6 shows a powerful main effect for targets which accounted for at least 15% of the variance. Individual target means were: .696, -.210, -.408, and -.106. All pairwise comparisons are significant ($p < .01$). The high score of target one, Alan Hill, on this variable lends support to the "Impulsivity-psychopathy" interpretation. No other main effects were significant.

Next to targets the most variance in impulsivity-psychopathy was accounted for by the target by label by source by suggestion interaction ($F(6,285) = 3.03$, $p = .007$). The sex by target by source interaction noted in Table 6 was also significant.

Component VI		
Threshold-Recidivism:	+.52 Threshold	+.44 Mature-youthful
	+.54 Recidivism	+.43 Stable-changeable
		+.42 Confidence

The sixth dependent variable reflects the recidivism judgments, as well as the threshold variable, reflecting the consistency of high psychopathology together with high recidivism. The semantic differential ratings reflecting lack of stability and maturity, also loaded in a consistent manner with recidivism and threshold on this variable. The confidence of the judges was also related to this variable.

Table 6 shows that only targets had a significant and major main effect on these threshold scores. The four targets means were: . .152, .232, -.154, and -.281. All pairwise comparisons were significantly ($p < .01$) different.

Table 6 shows that the target by source as well as the sex by target by source interactions were significant. Together they account for 4% of the recidivism variance. The sex variable was involved in another interaction (sex by label) which accounted for more than 2% of the variance in recidivism. All pairwise comparisons were significant at the .05 level. Males attributed less recidivism to targets in the patient label condition than in the normal background condition (M = -.266 vs. .113). The pattern for females was reversed (.228 vs. -.123).

Component VII		
Social Undesirability:	+.53 Warm-cold	-.61 Confidence
	+.51 Good-bad	-.43 Base rate
	+.47 Sensitive-insensitive	
	+.47 Mature-youthful	

The seventh dependent variable is an evaluative, or social desirability factor based primarily on the rating scale scores listed above: it appears to be a major determinant of confidence. Note that this factor is unrelated to the variable which measures the judge's estimate of the social desirability value for the target in the general population. However, threshold and judged base rate, measures of the judge's own evaluations rather than predictions about others, do load slightly on this factor. Thus, judges were least confident and gave lower base rate estimates when judging targets they rated cold, bad, insensitive, and youthful.

Targets again account for the most variance (20%). The means for the four targets were: .610, .184, -.099, and -.689 respectively.

All pairwise comparisons were significant beyond the .01 level. The target by suggestion interaction is the next most influential effect. The target means in the high suggestion condition were: .587, .115, -.175, and -.468 respectively, compared to .634, .254, -.023, and -.912 in the low suggestion condition. Within target comparisons show that all means were significantly different ($p < .01$) except Alan Hill. Note that the direction of the difference is reversed for Art Reynolds i.e., the high suggestion mean is more undesirable.

DISCUSSION

A reduction of the 19 original dependent variables by principle components decomposition produced a set of factors accounting for almost three quarters of the variance. The large, first factor which was interpreted as a mental illness variable, highly consistent with the first dimension in the preceeding study was clearly the predominant component. The remaining components were identified as excitability-impulsivity, prognosis-threshold, sensitivity, impulsivity-psychopathy, threshold-recidivism, and social undesirability. These components demonstrated considerable stability across methods of decomposition and rotation as well as an independent sample of judges.

Analyses of variance on these variables showed that various patterns of psychopathology, i.e., targets of varying prevalence, and/or diagnostic suggestions of severity were the most influential effects on these seven dependent variables in every case. This is consistent

with studies reporting major suggestion effects in the context of psychopathology (e.g., Di Nardo, 1975; Gordon, 1976; Lee & Temerlin, 1970; Temerlin, 1968, 1970; Yaffe & Mancuso, 1977) as well as education (e.g., Foley, 1977; Foster, 1976). In this study diagnostic suggestion was manipulated independently of source credibility, labels, target behaviors and other variables that have been confounding factors in important earlier studies (e.g., Di Nardo, 1975; Lee & Temerlin, 1970; Temerlin, 1968, 1970).

The relative weakness of labeling effects vis a vis targets sheds considerable light on the debate over whether labels or specific behavior patterns have a greater influence on judgments (e.g., Kirk, 1974, 1976). Targets, in other words, behavior patterns, dramatically affected six of the seven variables. Labels increased only the prognosis scores; they did not lower sensitivity, as was the case in Paquin and Jackson (1977), or mental illness, the largest dependent variable, or in fact, any of the other dependent variables in this study. On the other hand, labels were involved in many interactions for most of the dependent variables.

A similar story unfolds for the source credibility variable, which like labels had a significant impact on prognosis but no other variable; The highest prognosis scores were found in the secretary-clerk condition, the lowest for the social worker! Source credibility, like labels, was involved in numerous interactions. Likewise, sex of judge often interacted with other factors in complex ways. However, sex produced no main effects.

One explanation for the presence of numerous interactions may

lie in what Rosenhan (1975) calls context effects in psychodiagnosis. In the present study the context of a judgment refers to the differential stimulus conditions of the task for a given judge and is represented by the observed interaction effects. The target by suggestion interaction provides a good example of this. In all but two of the dependent variables their was a differential effect of diagnostic suggestion related to the type of target person, or type of psychopathology being judged. It appeared that diagnostic suggestions were most influential when they were consistent with behavior, i.e., in targets with lower prevalence and higher judged pathology. An important point to keep in mind with respect to these results and an important implication for clinicians is the unpredictability of most of these context (interaction) effects across relevant judgmental dimensions. With few exceptions the same stimulus or judgmental context was not involved in significant interactions across the same set of variables. To further complicate matters, characteristics of the judges (sex) often interacted with judgmental contexts, and again, in an unpredictable manner vis à vis the seven, orthogonal dependent measures of this study.

Mental illness, the most important of these variables, proved to be extremely sensitive to contextual and subject differences. The significant four and five way interactions are noteworthy in this regard. The prognosis-threshold component was similarily influenced by complex variations in context and judges although not to the same context variations as mental illness. It is important to note that labels had a significant impact only upon this prognosis variable but

furthermore, that the patient label resulted in better prognosis and higher threshold scores! The apparently beneficial effect of labels may be attributable to the expectations produced in judges by the label-suggestion context they experienced. It seems safe to conclude that mental illness and prognosis are complexly determined and sensitive to many of the judgmental variations that exist across clinical judgment contexts. Unawareness of these contextual influences may not only reduce the accuracy of our predictions it may cause them to be in the wrong direction. Such was the case in the observed beneficial effects of labels on prognosis judgments.

The excitable-impulsive, impulsivity-psychopathy, and social desirability dependent variable components, on the other hand, were influenced by a relatively small number of factors. The social desirability variable, for example, was influenced by targets and the target by suggestion factor, as were most of the dependent variables. Aside from these factors, only the effect reflecting sex of judge in combination with specific labels and suggestions was significant in modifying social desirability scores. The two variables tapping aspects of an impulsivity dimension were both significantly affected by the target by label by source by suggestion interaction as well as by diagnostic suggestion in one case and targets in the other. While the influences on these impulsivity dimensions are relatively few they are indeed both large and complex.

Targets were the most important factors in threshold-recidivism scores. Consistent with previous variables (i.e., mental illness and sensitivity) the most prevalent pathological target received extreme

scores; John Walker, the most prevalent pathological target was judged most likely to have recurring problems and high levels of psychopathology. The second largest influence on threshold was the sex by label interaction; females attributed more pathology and recidivism to targets labeled as ex-psychiatric patients but male judges responded the opposite way. Similar findings were observed for the target by source context; males and females responded differentially to this context suggesting that sex of judge may interact with the credibility (or prestige) status of the source of a clinical, assessment report when judging patients of varying profiles.

Finally, a number of important findings were revealed for sensitivity, the component of interential accuracy which reflects the judge's ability to accurately predict the target's (e.g., a patient's) responses. As expected judge's were most sensitive to the most prevalent target supporting an assumption of implicit personality theory, namely that the accuracy of our implicit theories about others derives from our social contacts and cultural experiences with them. Judges were least sensitive to Art Reynolds, perhaps because they could not believe he was so problem-free in the context of rating three other targets which were so obviously psychopathological. The reader may have noted the similarity of this scenario to the one reported by Rosenhan (1973) who concluded that current diagnostic procedures could not differentiate the sane from the insane. An alternative explanation is that varying label and suggestion contexts may have seemed more inconsistent with Art Reynolds relative to the other targets thereby making judgments about Reynolds more difficult.

Little support for the latter can be obtained by examining the significant interactions for sensitivity. Judges were more sensitive to Reynolds in the high suggestion condition, a finding which is more supportive of the "too good to be true" hypothesis. The significantly higher mean for Reynolds in the high suggestion, patient label condition lends further support to this. This suggests that perhaps clinicians need to be aware that judgmental context may play a major role when patients vary in pathological severity and prevalence (i.e., familiarity to the judge).

There are numerous sources of bias in clinical judgment most of which detract from rather than enhance accuracy. One advantage of this study is the independent manipulation of several of these important variables i.e., targets (behavior patterns), labels, diagnostic suggestion, source credibility, and sex of judge. While the investigation of a large number of variables creates a complex design yielding results that are difficult to interpret, it also permits the study of complex interactions which reflect what has been called the diagnostic or judgmental context. The use of measures based on the model of inferential accuracy (Jackson, 1972) to assess judgmental accuracy represents an improvement relative to the complete reliance of most other studies on rating scales (Abramowitz & Dokechi, 1977).

In summary, this study found that behaviors were more influential overall than labels. This must be qualified however, by noting that labels did have a substantial effect on prognosis scores but unexpectedly in a positive direction. This was attributed to the effect of task and stimulus context on the expectations and responses of the judges. Furthermore, labels, in combination with other variables

often did produce significant effects; the same is true of sex of judge. It was this presence of numerous interaction effects which supported the notion of the existence of complex, diagnostic, context factors in these clinical judgments.

To conclude, there is little doubt that (target) behavior patterns were the most influential factors across the dimension of judgments in this study. This holds true even when the non-pathological target, which produced the biggest effects, is ignored. Therefore, it is reasonable to conclude that the prevalence of the target, a measure of the judge's familiarity with that pattern of psychopathology, may be as important as content in clinical decision making about patients. Diagnostic suggestions about severity, of the type often found in assessment reports and referrals, were also very influential. On the other hand the credibility or prestige of the source of the diagnostic suggestion did not, with the exception of prognosis judgments, produce direct effects, even though in conjunction with other variables it did affect judgments. Perhaps more important is the finding that judgmental contexts, such as target by suggestion interactions, also affect accuracy suggesting that further studies of complex biases in clinical judgment are still needed. One important area for research, not addressed in this study, which is suggested by these findings concerns the issue of patient-therapist matching (cf. Berzins, Ross & Cohen, 1970; Persons & Marks, 1970; Staples, Sloane, Whipple, Cristol, & Yorkston, 1976). It is possible that patient characteristics such as sex, social class, age, insight, verbal ability etc., may interact with characteristics of the diagnostician and therapist to produce judgmental errors that have immediate

consequences for both diagnosis and treatment. For this reason the importance of the total judgmental and clinical context cannot be overstated, and it should serve both as a warning beacon for clinicians and a guidepost for researchers.

CHAPTER 4

Information inconsistency vs. cognitive focus in labeling effects

The purpose of the present study is the exploration of the processes involved in the labeling effects observed in various studies of clinical judgments. Using the model of inferential accuracy (cf. Jackson, 1972; Reed & Jackson, 1975) Paquin and Jackson (1977) found that labels did bias judgments about certain patterns of psychopathology. They suggested that the labeling effects might be accounted for in two ways. In the first place specific labels may provide information which judges perceive as incongruent with additional psychopathology or trait information. This discrepant information causes difficulty in synthesizing disparate data in the judgment task resulting in lowered levels of consensus and prediction. The phenomenon is termed "trait-label inconsistency," and is concordant with previous studies citing the importance of information consistency variables in a variety of judgment contexts (e.g., Gross & Crofton, 1977; London & Hakel, 1974; Manis & Moore, 1978).

The second,related explanation for the labeling effects observed in Paquin and Jackson (1977) involvesan information salience hypothesis. It was suggested that labels may create a judgmental set to focus more upon the labels themselves, perhaps because of their information summarizing capacity, rather than upon the network of traits and specific behaviors implied by the personality description in which the label was imbeded. This phenomenon is termed "judgmental focusing." Rosenhan (1975) in discussing context effects in psychiatric

judgments, describes a similar process, whereby diagnoses are influenced by expectations and the environment in which the judgments are made, particularly when there is ambiguity in the situation. Thus, recall and perception of stimuli often become selectively focused and as Rosenhan notes "These powerful effects occur because neither memory nor perception are passive processes" (p. 464).

METHOD

Design Overview and Purpose

The intent of this study is to examine the potential contribution of the Trait-Label Inconsistency hypothesis and the Judgmental Focus hypothesis in explaining labeling effects in clinical judgments. Following the paradigm employed by Paquin and Jackson (1977) subjects read brief psychological descriptions of actual, empirically identified, patient types and then were asked to predict the likelihood that the described person would endorse a set of DPI items. Each subject judged four such "target" descriptions which varied in level of trait-label inconsistency. Trait-label inconsistency is defined in terms of projections in the multidimensional space previously generated (Chapter 2). In order to test the judgmental focus hypothesis some subjects were instructed to focus their attention on current, specific information contained within the description (i.e., trait information) and to weight this information highest. Other subjects were told to focus on summary descriptions and previous history data (i.e., labels). A third

(control) group were given no focusing instructions. As a test of the effectiveness of these instructions subjects were later asked to rate in terms of judgmental importance, all the information bits contained in the descriptions.

Judgments of the descriptions were evaluated in terms of Jackson's (1972) model of inferential accuracy as in the preceeding study. The model has two parameters: (i) Threshold, is the generalized tendency to attribute psychopathology and can be defined as the subjects' mean rating when rating scales are used in the judgment task; (ii) Sensitivity is the awareness of trait inter-relatedness or behavior patterns and can be operationalized as the correlation between a subject's predictions of a target's item responses and the target's actual responses.

Thus, the design is a 4 X 2 X 3 X 2 (targets X consistency X focus X sex of judge) fixed factor, repeated measures (targets) design with subjects nested in the label consistency and instructed focus factors. In addition to the primary dependent variables, threshold and sensitivity, selected prognostic judgment ratings are also analyzed as in Chapter 3 (see Appendix B).

Subjects

Judges for this study were volunteers from undergraduate psychology courses who received credit for their participation. There were 189 judges; 79 were male and 110 were females.

Procedure

Subjects were told that this was a study of clinical judgment processes designed to assess the feasibility of a recently developed approach to the classification of psychopathology for the improvement of accuracy in clinical judgment tasks. Judges were given four (target) descriptions of psychopathology and were instructed to read the descriptions carefully before rating the probability that the target person would endorse the set of 52 DPI items. They also responded to the set of 17 clinical and prognostic judgment items for each of the targets (appendix B).

Subjects received target descriptions which contained labels which were either consistent or inconsistent with the target, as described below. In addition some received an instructional set which focussed their attention either on "summary sources of information and background information" (labels) or, on "specific traits and examples of behavior" (target descriptions). After judging each target they were asked to identify the informational bits contained in the descriptions which they felt were most influential. This was done in order to assess the effectiveness of these judgmental focusing instructions (i.e., a manipulation check). These data were scored in terms of the number of labeling (vs. trait) information units selected by the judge.

Judgmental Focusing Instruction

These instructions were designed to provide a rationale for,

and a cognitive set either to focus on the labeling or the trait information in the judgmental task. All judges were told that psychological assessment reports vary in the nature and specificity of the information they contain and that the reports (target descriptions) which they would judge were descriptions of actual individuals who had "undergone a routine psychological assessment as part of a large company's employee benefits program." All judges were asked to: "(1) First, read the summary assessment and associate the main characteristics with the name of the person (testee) involved. (2) Then read it again concentrating your attention on the nature of the information contained in it ..." They were also all given instructions on how to make their judgments and use the answer sheets.

Subjects in the No-Focus Condition received no further focusing instructions. Subjects in the Label Focus Condition were told that one of the most useful types of assessment information for such judgments are "summary concepts" since they reflect multiple sources of information within a single, brief, and easy to recall term of phrase. Thus, they were instructed to concentrate their attention on summary sources of psychological information such as sex, diagnosis, and previous history when available:

> "When making your judgments about the described person, therefore, focus your attention on these summary concepts and weight these information sources highest when deciding upon your answers (ratings)."

The actual instructions for subjects in the label focus condition are given in Appendix C. The instructions for the trait-focus condition were identical to these except as described below.

Subjects in the Trait-Focus Condition were instructed to concentrate their attention on specific incidents or examples of behavior rather than vague terms of diagnostic labels:

> "Please focus your attention upon the specific traits and explicit behaviors contained in the report since such information increases accuracy and the validity of clinical judgment tasks such as this one; you should therefore rely heavily on this kind of information. Some examples of this trait and behavioral information are: specific feelings or emotions (e.g., 'has been feeling sad and depressed') or incidents of behavior (e.g., 'has consulted his doctor often' or 'felt uncomfortable talking with his friends'). The information segments most like 'specific traits and explicit behaviors' are labeled D to I on the assessment reports. Thus, your judgments about the described person (testee) should be most influenced by the D to I segments."

Trait-Label Consistency

The target descriptions, which contained the psychopathology trait configuration, and their corresponding labels were selected from study one (Chapter 2) on the basis of the availability of labels at the two levels of consistency for each target. Targets were selected which had strong projections on one of the dimensions identified in study one. Target projections on these dimensions and the selected labels at the two levels of consistency are described in Table 1. Consistent labels are defined by a moderate to high projection in the same direction and on the same dimension as the target description, whereas inconsistent labels had opposite projection coefficients. The actual target descriptions are presented in Appendix A. Thus, subjects were randomly assigned to one of two trait-label consistency conditions in which either an inconsistent or consistent label was paired with the target description.

Table 1

Characteristics of the Four Target Descriptions and Labels

	Target	Modal[a] Profile	Defining[b] DPI Scales	Projection on MDS Dimensions[c]					Prevalence[d]
				I	II	III	IV	V	
1.	Jack Cole	2+	Desocialization, Ideas of Persecution, Sadism, Shallow Affect	-.14	.93	-1.11	-1.66	-.39	101 (10.7%)
2.	John Walker	1+	Health Concern, Repression, Shallow Affect	.46	-1.01	.01	.75	-.16	239 (25.3%)
3.	Peter Carling	6-	Hypochondriasis, Somatic Complaints, Perceptual Distortion, Shallow Affect	-.14	-1.32	.18	.15	.08	20 (2.1%)
4.	Art Reynolds	Normal	None	3.97	-.77	-.13	.04	-.27	----------
	Labels								
	Depressive neurosis			.02	.27	1.48	.39	-.16	
	Ex-convict			.42	1.16	-.92	-.81	.04	
	Alcoholism, episodic excessive drinking			.54	1.16	.21	.07	-.41	
	Schizophrenia, simple type			-1.25	-.85	-.27	-.25	-.20	

[a] The empirically identified profile type which defined the targets; from Reed (1976).

[b] The Differential Personality Inventory scales which defined the modal profile configuration.

[c] The projections of stimuli on the multidimensional scaling solution reported in study one.

[d] The number of people classified by the given profile in the (N = 953) population of Skinner, Reed, and Jackson (1976).

It should be noted that in order to control for the potential confounding of target-label consistency with the denotative meaning and idiosyncratic characteristics of the labels, pairs of targets were crossed (counterbalanced) with labels. Thus, each target was paired with the same labels as the one other target having the opposite target-label consistency configuration. The Art Reynolds and Peter Carling targets, for example, are both paired with the labels "schizophrenia, simple type" and "alcoholism, episodic excessive drinking." However, these labels are related in opposite ways to the targets. The target-label consistency manipulation can be summarized as follows:

Target	Inconsistent Label	Consistent Label
Jack Cole	depressive neurosis	ex-convict
John Walker	ex-convict	depressive neurosis
Peter Carling	alcoholism, episodic excessive drinking	schizophrenia, simple type
Art Reynolds	schizophrenia, simple type	alcoholism, episodic excessive drinking

The targets, as in the preceeding study were described in the context of a clinical assessment report. However, all reports were allegedly written by the same, moderate credibility source (a MSW) and described the history of the target as a patient released from a psychiatric hospital in 1970. Representative examples of these clinical reports are included in Appendix C.

RESULTS

The dependent variables and the statistical analyses are identical to those of the preceeding study. A principle components solution was employed to decompose (i.e., condense) the dependent variables. The coefficients of congruence between the first seven varimax rotated factors and the corresponding factors in study two following orthogonal procrustean rotation using the study three matrix as a criterion were: .93, .95, .84, .96, .94, .94, and .87. Given the great similarity of the two matrices, the seven mutually orthogonal component variables of this study are interpreted in the same manner described in the preceeding study, and component (or factor) scores derived from these seven factors were computed for each subject. These seven dependent variables were subjected to analyses of variance (Table 2). The outcome of tests of the assumptions of the normality and homogeneity of variance of the data was virtually identical to that of study two. Post hoc tests reported below employed the Duncan new multiple rating procedure using the .01 level for alpha unless otherwise indicated.

The manipulation check revealed that the focussing instructions had the intended effect of modifying judges' cognitive set about which information bits were most important for the judgments. According to responses to the question "What information segment(s) influenced your overall judgments about this person?" judges in the label focus condition reported being influenced by more labeling information (M=1.13, s.d.= 2.8) than the judges in the trait focus (M=.65, s.d.= 2.6) or no

Table 2: ANOVA Summary for Three Dependent Variables

Effect	df	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS
		Mental Illness				Excitable/Impulsive				Prognosis/Threshold			
G	1,177									7.49	12.0**	.009	.048
LF	2,177									2.39	3.8*	.006	.012
S	177	1.36				1.21				.62			
T	3,531	66.36	129.1**	.270	.904	6.38	7.5**	.024	.202	125.88	312.7**	.512	.878
TL	3,531					5.83	6.8**	.022	.182	1.84	4.5*	.007	.036
TF	6,531					1.69	1.9	.012	.031				
GTF	6,531					1.80	2.1	.013	.035	11.49	2.8*	.009	.005
TLF	6,531	1.05	2.1	.008	.007	1.78	2.0	.013	.034				
ST	531	.51				.85				.40			

* $p < .01$.

** $p < .001$.

NOTE: All non-significant ($p > .05$) effects have been omitted. $\hat{\omega}^2$ MS is an alternative measure of explained variance ($\hat{\omega}^2$), based on $\{MS_i - MS_e\} \div \Sigma MS$. G = Gender (sex of judge), L = Label Consistency, F = Focus, S = Subjects, T = Targets

focus condition (M= .59, s.d.= .80), $F(2,177) = 6.04$, $p = .003$. Further, it was unlikely that these judges were reponding indiscriminantly or to a demand characteristic since there was no significant difference ($p = .25$) in the total number of influential, information bits identified by judges in the three conditions.

Table 2 shows that mental illness, the largest component variable, was almost entirely determined by the highly significant ($F(3,531) = 129.1$, $p < .0001$) target effect which accounted for 27% of the estimated population variance and 90% of the experimental variance. The means for the Jack Cole, John Walker, Peter Carling, and Art Reynolds targets were: .26, -.37, .74, and -.64 respectively. Hereafter, target means are reported in the same order. All pairwise comparisons among targets were significant. The target by label by focus effect was also significant but accounted for less than 1% of the mental illness variance.

The pattern of effects for the excitable-impulsive variable is more complex. Again targets were highly significant and accounted for the most variance (see Table 2), although considerably less than for mental illness. The target means were: .15, -.16, -.18, and .17 respectively. The first (Cole) and fourth (Reynolds) targets each were significantly different from the remaining targets.

The target by label interaction was equally important. The target means in the consistent label condition were: -.01, -.11, -.41, and .34 compared to .31, -.21, .05, and .01 in the inconsistent label condition. Within target comparisons revealed that all means were significant except for John Walker. For the latter however, the

inconsistent label mean was significantly lower than the consistent mean, at the .05 level. Thus, the direction of the difference is the same as the Art Reynolds target. Note however, that for the Jack Cole and Peter Carling targets, the least prevalent of the three pathological targets, the inconsistent label condition produced higher means on excitability-impulsivity.

Targets also interacted with the focus condition although this effect accounted for less than 2% of the excitability-impulsivity variance. The most important finding revealed by pairwise comparisons concerns the no-focus (control) condition, which produced the lowest means (though not significantly different from the label focus) on the least prevalent targets but higher means on the other two targets, Art Reynolds in particular. In fact, while the label vs. trait focus means were identical (both .03) for Art Reynolds, the no focus mean was .47, significantly higher than any other mean for this effect.

Table 2 shows that the prognosis-threshold variable is most influenced by targets, which accounted for more than half the variance. John Walker and Peter Carling had the highest prognosis scores (.73 vs. .67) and did not differ significantly. Jack Cole (-.58) and Art Reynolds (-.91) were significantly lower than the other targets and significantly different from each other. Sex of judge was another significant, though relatively small main effect with males having significantly lower scores than females (-.12 vs. .08) respectively.

Pairwise comparisons of the label by focus interaction revealed that label consistency means were not different in the label focus condition but the consistent label means were higher in the trait

focus condition while lower than the no-focus condition. The label by target and sex by target by focus interactions were also significant but small and difficult to interpret.

The sensitivity variable (not reported in the tables) did not yield any significant effects except for targets ($F(3,531) = 46.1$, $p < .0001$, which accounted for more than 12% of the estimated population variance or 61% of the (experimental) variance according to $\hat{\omega}^2$MS. Judges generated the highest sensitivity scores for John Walker (.42), the most prevalent pathological target, followed by Jack Cole (.30) the second most prevalent target, although the difference is not significant. The means for Peter Carling (-.38) and Art Reynold (-.36) obviously do not differ from each other, but both are significantly lower than John Walker and Jack Cole sensitivities.

Main effects due to targets and sex, which explained about 17% and 1% of the impulsivity-psychopathy variance respectively, are noted in Table 3. Not surprisingly, pairwise comparisons revealed that Jack Cole, the psychopathic target, was significantly higher than all other targets. No other target comparisons were significant. Males gave significantly lower impulsivity scores (-.13) across all targets than did females (.09).

The target by label interaction explains even more of the variance than the main effect for sex of judge. The means for the targets in the consistent label condition were: .82, -.36, -.37, and -.36 vs. .58, .00, -.09, and -.38 in the inconsistent label condition, respectively. All within target comparisons were significant except Art Reynolds. Furthermore, both Jack Cole means were significantly

Table 3: ANOVA Summary for Three Dependent Variables

Effects	df	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS	MS	F	$\hat{\omega}^2$	$\hat{\omega}^2$MS
		Impulsivity/Psychopathy				Threshold/Recidivism				Social Undesirability			
G	1,177	8.44	5.2*	.009	.111					9.53	8.9*	.012	.076
GL	1,177									6.29	5.9	.007	.047
LF	2,177									3.95	3.7	.009	.026
S	177	1.60				1.55				1.06			
T	3,531	41.81	72.9**	.166	.669	15.29	2.09**	.060	.343	83.81	187.3**	.344	.743
GT	3,531					2.63	3.6 *	.010	.045				
TL	3,531	3.28	5.7**	.012	.044					1.71	3.8*	.006	.011
TLF	6,531					1.56	2.1	.011	.020				
ST	531	.57				.73				.45			

* $p < .01$.

** $p < .001$.

NOTE: All non-significant ($p > .05$) effects have been omitted. $\hat{\omega}^2$MS is an alternative measure of explained variance ($\hat{\omega}^2$), based on $\{MS_i - MS_e\} \div \Sigma MS$. G = Gender (sex of judge), L = Label Consistency, F = Focus S = Subjects, T= Targets.

higher than all other means; note that for this target, but no other one, the consistent label mean produces higher impulsivity-psychopathy scores. In other words, label consistency may indeed produce differential effects, depending on the similarity of denotative meanings of the label and target and on the relevance of the judgmental dimension being used.

Table 3 reveals that the main effect for targets explains the most threshold-recidivism variance followed by the target by label by focus interaction. The four target means were: .22, -.05, .20, and -.41. The Art Reynolds and John Walker targets each were significantly lower than all other means suggesting that recidivism based threshold scores may depend upon the frequency of occurrence of the target. There was a trend ($F(2,177) = 3.91$, $p = .08$) for judgmental focus observed, in the predicted direction, for threshold. The label, trait and no-focus means were: .13, -.04, and -.12.

The social undesirability variable showed significant main effects for targets, by far the most important effect, and sex (males > females). The target means were: .97, -.02, -.22, and -.66; All pairwise comparisons were significant. The lack of a main effect for label consistency may, in part, be explained by the remaining effects of significance. The sex by label, focus by label and target by label interactions were all significant, suggesting that label consistency effects on desirability scores depend upon the judgmental context and characteristics of the judge.

DISCUSSION

One thing is clear. Across all dimensions of clinical judgment in this study the characteristics of the person being evaluated or judged are the most influential. For some dimensions the psychopathological content or severity appeared to be the relevant target characteristic. For other dimensions e.g., mental illness, sensitivity, and threshold-recidivism, the prevalence of the target, reflective of the extent to which judges are familiar with target characteristics, appears to be relevant. This would be an important aspect of target characteristics to directly test in future research.

If labels have an impact on clinical judgments of the sort measured in this study, it is not likely that the trait, or behavior-label consistency is the element which is directly responsible. However, trait-label consistency does influence judgments to some extent, particularly social undesirability, in the context of specific patterns of psychopathology and judgmental sets. This is also illustrated by the target by label by focus interactions on mental illness, excitable-impulsive, and threshold-recidivism scores.

The focusing instructions, while apparently successful in creating appropriate cognitive sets in the judges to use either labeling or behavioral information, were not successful in increasing accuracy or modifying judgments. There was, however, a trend noted for the focus effect on the threshold-recidivism variable in the predicted direction; judges in the trait focus condition rated less recidivism and psychopathology than the label focus judges, but more than judges in the no-focus condition. Instructing judges to focus on behaviors rather than labels may yet prove to be significantly effective in lowering

psychopathology judgments related to recidivism. However, would this mean more accurate judgments? Judgmental focus did not affect sensitivity in this study. Furthermore, even lower scores tended to be produced in the no-focus condition! Obviously much more intensive background experiences (Langer & Abelson, 1974) and specific training programs (e.g., Strasburger & Jackson, 1977), as well as evaluative research, are necessary in order to teach judges unbiased, valid, cognitive sets resulting in greater accuracy in clinical judgments. It appears that such training might generalize across sex of judge since there were few interactions involving sex and focus or label consistency. However, more research on this and other aspects of improving clinical judgments is still needed.

In conclusion, neither the trait-label inconsistency nor the judgmental focusing hypothesis received direct support as simple and major explanations for the processes underlying labeling biases observed in the literature. These processes did consistently interact with each other as well as targets suggesting that they may play a role in the total judgmental context. The importance of judgmental contexts, noted in the preceeding study, is re-affirmed here. Recall, for example, the differential effects of label consistency as well as judgmental focus on excitability-impulsivity judgments depending upon the familiarity of judges with the target. Such differential effects may also occur when making judgments of relatively normal targets, a finding which was observed for impulsivity-psychopathy judgments. Clinicians and researchers should take note therefore that label consistency and judgmental focus effects may vary with target characteristics, and with the judgmental criterion as well.

If consistency and focus (cognitive set) cannot fully explain alleged labeling effects what are some other possible factors? Perhaps judges prefer to rely on brief, discrete sources of information. Perhaps labels serve as an information summarizing mechanism. Perhaps even, labels contain as much valid information as they do information with only an illusory correlation (Chapman & Chapman, 1967, 1969) to the criterion. It is difficult to isolate the valid and reliable information, if any, inherent in a label, from "stereotypical" or intuitively logical though false information. For example, even the label "former psychiatric patient" may contain useful base rate information; even though a judge cannot be certain that characteristics of this prior status are currently relevant there is some probability that being a former patient conveys valid information (e.g., susceptibility to stress or the judgments of others). Reseach has yet to demonstrate whether this probability fluctuates with the judgmental context, type of judge and target person, and the situation.

The logical consequence of this is to conclude that labeling information may enhance accuracy in some contexts. Perhaps the conclusion that labels are an illusory correlation is itself an illusion based upon methodological artifact. One illustration of such an artifact is the use of analog studies using judgment tasks lacking the complexity of actual clinical judgments. It appears that labels do indeed exert powerful effects when they are highly salient (e.g., Paquin & Jackson, 1977). However, when labels are embedded in a larger judgmental context the strength of the bias may be diminished.

While the present study was not designed to test the effect

of labels on accuracy and other aspects of clinical judgments, evidence presented in the preceeding study suggests that labels do not directly influence judgments. The present findings of powerful and pervasive target (behavior) effects lend support to this conclusion.

CHAPTER FIVE

Summary and conclusions

The results of the multidimensional scaling study supported the hypothesis that a relatively small number of dimensions can account for judgmental dimensions of psychopathology. The five interpreted dimensions, which demonstrated notable levels of stability, explained 95% of the judgmental variance, with the first dimension, mental illness, accounting for half of the total variance. Factor analyses of the unidimensional ratings confirmed the interpretation of the five dimensions as mental illness, social vs. psychological deviance, depression vs. impulse control, anxiety vs. impulse control, and paranoia vs. impulse control. Of greater importance is the observation that the factoring of the dependent variables in the other studies revealed a notable similarity, particularly in the case of mental illness, to the dimensions of study one. Since the stimuli varied considerably, this would suggest that the obtained dimensions generalize well across varying sets of psychopathology descriptors. The similarity of several of these dimensions to those reported by Chan and Jackson (1978) provides further support for the generalizability of these dimensions.

In conclusion, it appears to be the case that few dimensions are needed to account for judgments of psychopathology whether based upon lay labels or allegedly more precise nosological terms represented by the DSM II. It is not clear, as yet, whether the new taxonomy, the DSM III, will represent an improvement in this undesirable state of affairs (e.g., Schact & Nathan, 1977). Qualifying phrases incorporated

into DSM II diagnoses had minimal beneficial impact on judgments. There was, however, some evidence that the modal profiles, specific behavioral exemplars of psychopathology, may be a promising alternative to the classification of psychopathology because of their informational specificity.

The second study examined the independent effects of target behavior, diagnostic suggestion, source credibility, labels, and sex of judge on accuracy and other dimensions of clinical judgments. Original dependent variables were factored into a set of variables representing seven dimensions, which showed great stability and generality. These seven variables, which were replicated in study three on an independent sample, were: mental illness, excitability-impulsivity, prognosis-threshold, sensitivity, impulsivity-psychopathy, threshold-recidivism, and social undesirability.

Target behaviors and diagnostic suggestions exerted powerful effects on almost all seven variables. It was concluded that behaviors rather than labels and diagnostic suggestions rather than (and virtually independent of) source credibility or prestige were clearly the most important direct influences on these major dimensions of judgment. While this study employed the label "former psychiatric patient" evidence from the first study, and in the literature (e.g., Berman, 1976) suggests that judges do not differentiate patient labels on the basis of prior vs. current status. On the other hand, it was suggested that the interaction of labels and source credibility with these direct influences supports a contextual interpretation of psychodiagnosis and clinical judgments as has been argued by Rosenhan (1975) and others

(i.e., Estell, 1975; Fitzgibbons & Shearn, 1972; Weinberg, 1977). It was also suggested that characteristics of the judge, such as sex, may interact with these judgmental contexts thereby further complicating our understanding of clinical judgment.

The third and final study used a similar design and the identical analyses in order to investigate alternative explanations for labeling effects. Neither the consistency of the label relative to the target's behavior nor the successfully manipulated, cognitive focusing instructions generated consistent or powerful results. It was suggested, however, that such explanations may be involved in the larger judgmental context and in this sense the importance of context effects noted in the second study was re-affirmed here. Future research may reveal that styles of information processing such as scanning vs. focusing may vary across judges and be differentially appropriate in changing judgmental contexts.

It is concluded that the importance of context effects cannot be overemphasized for two reasons. First, they offer an approach to conceptualizing the conflicting results observed in the literature with regard to the effects of various sources of bias on clinical judgments. Second, judgmental context effects have direct implications for suggestions about the improvement of clinical inferences and decision making. For example, it was observed that on some dimensions of judgment, the judges' degree of familiarity with the pattern of psychopathology appeared to influence the extent to which various judgmental contexts modified accuracy, the propensity to perceive psychopathology, and psychological attributions to the target.

Training background and clinical experience is related to the tendency to be influenced by certain judgmental biases (e.g., Langer & Abelson, 1974). This suggests that once the precise nature and the mechanics by which biases operate have been identified, appropriate methods can be devised to improve the accuracy of clinical inference, diagnosis, and important decision making related to clinical judgments affecting treatment. The studies in this thesis have contributed to this progress by beginning to map the major perceived dimensions of psychopathology judgments of clinical relevance, by assessing the relative importance of various influences on clinical judgments, and by identifying how biases probably operate in specific judgment contexts modelled after actual clinical activities in psychiatric settings. The generality of these findings is enhanced by studies showing that judgmental accuracy and susceptibility to bias may not be different across diverse groups such as clinicians, students and workers (e.g., Goldberg, 1959, 1968; Simmons, 1976).

In the past studies have generally failed to improve the judgmental abilities of clinicians, in part because of the reliance on experimentally pragmatic but clinically irrelevant and unfair tasks, and in part because of ignorance about the major impediments to accuracy and the corresponding optimal training techniques. A few recent studies (Cohen, 1976; Lindel, 1975; Strasburger & Jackson, 1977), by identifying some useful kinds of information, rekindle optimism in the feasibility of improving accuracy in meaningfull clinical judgments by means of a feedback-learning paradigm. Armed with such an approach and knowledge of specific biases presented here, research and training can now move further toward overcoming biases and improving accuracy in clinical judgments.

The present studies point to information about a target person's behavior as well as suggestions regarding the severity or frequency of that behavior as very important elements for such research and training efforts. It was noted that other variables such as labels and source credibility did play a role within the larger judgmental or informational context. It is possible that labels and source credibility may have produced main effects had they been manipulated differently. For example, the provision of detailed introductory information about the experience, training, credentials etc., of the source may influence judgments about a source such as a therapist or counselor (Scheid, 1976). Future research may reveal that introductory information about the credibility and prestige of the source of judgment information may in fact also modify judgments about the patient or client involved. Study three of this thesis did not find that source credibility, which was embedded in the context of powerful target behavior and diagnostic suggestion information, modified judgments of the described person. Perhaps the use of implicit conceptions about psychopathology mitigates against the direct influence of source credibility on clinical judgments about a target. In other words, when inferentially relevant information, such as the behavior pattern of the target, is presented to judges this information is given most judgmental weight and other sources of information, such as source prestige or previous diagnoses, become less important and perhaps even less salient to the judge.

The pervasive influence of target behavior in the present studies suggest that judges may utilize information which is most relevant to their implicit theories of personality and psychopathology. Assuming that such theories are valid, this would imply that judges

employ a sound strategy in their clinical judgments. Indeed recent research (Chan & Jackson, 1978) suggests that implicit theories of psychopathology are basically veridical.

Furthermore, one possibility which the ineffectiveness of the target-label inconsistency manipulation presents is that judges, quite properly, utilized the specific behavioral information even when presented with inconsistent labeling information which might have modified their judgments (and even when instructed to utilize such labels). Alternatively, one could posit that labels do not possess valid information in addition to the (pejorative) evaluative components they contain and that therefore there was no information inconsistency relative to the valid information contained in and implied by the target descriptions. In short, judges in the present study responded appropriately, probably according to their veridical implicit theories of psychopathology. These "theories," or conceptions appeared to be relatively resistant to labeling and source credibility biases.

The present results diverge from many previous findings with respect to the relative impact of labels and target behavior, and with respect to implications for conclusions about the adequacy of clinical judgments. It was argued that improved task meaningfulness, as described in chapter one, accounts for the largely appropriate responses of the judges with respect to predictions about psychopathology. Clinical judgments of an evaluative nature were facilitated by the use of specific, straightforward, relevant ratings. The investigation of several variables simultaneously, creating judgmental contexts similar to actual clinical judgments, was postulated as one improvement on previous research which can explain the absence of labeling effects and other

discrepancies of the present findings with previous research. The use of less global measures of clinical judgment, and an external criterion of accuracy based on the model of inferential accuracy are additional, advantageous differences of the present approach. This approach offers the researcher a useful basis for the continued study and improvement of clinical judgment.

REFERENCES

Abramowitz, C. V., & Dokecki, P. R. The politics of clinical judgment: Early empirical returns. Psychological Bulletin, 1977, 84, 460-476.

Adinolfi, A. A. Relevance of person perception research to clinical psychology. Journal of Consulting and Clinical Psychology, 1971, 37, 167-176.

Aloia, G. F. Effects of physical stigmata and labels on judgments of subnormality by preservice teachers. Mental Retardation, 1975, 13, 17-21.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders (2nd ed.). Washington, D. C.: Author, 1968.

Arthur, A. Z. Diagnotic testing and the new alternatives. Psychological Bulletin, 1969, 72, 183-192.

Berman, D. S. Judgmental bias and perceived mental illness (Doctoral Dissertation, Kansas State University, 1975), Dissertation Abstracts International, 1975, 36, 2517-2518B (Order No. 75-25, 019, 73).

Berzins, J. I., Ross, W. F., & Cohen, D. I. Relation of the A-B distinction and trust-distrust sets to addict patients' self-disclosures in brief interviews. Journal of Consulting and Clinical Psychology, 1970, 34, 289-296.

Bieri, J., Atkins, A. L., Briar, S., Leaman, R. L., Miller, A., & Tripodi, T. Clinical and social judgment: The discrimination of behavioral information. New York: John Wiley & Sons, 1966.

Billingsley, D. Sex bias in psychotherapy: An examination of the effects of client sex, client pathology, and therapist sex on treatment planning. Journal of Consulting and Clinical Psychology, 1977, 45, 250-256.

Calhoun, L. G., Pierce, J. R., Walters, S., & Dawes, A. S. Determinants of social rejection for help seeking: Locus of causal attribution, help source, and the "mental illness" label. Journal of Consulting and Clinical Psychology, 1974, 42, 618.

Carmer, S. G., & Swanson, M. R. An evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. Journal of the American Statistical Association, 1973, 68, 66-74.

Cauthen, N. R., Robinson, I. E., & Kraus, H. H. Stereotypes: A review of the literature, 1926-1968. Journal of Social Psychology, 1971, 84, 103-125.

Chan. D. W., & Jackson, D. N. Implicit theory of psychopathology. Multivariate Behavioral Research, 1978, in press.

Chapman, L. J., & Chapman, J. P. Genesis of popular but erroneous psychodiagnostic observations. Journal of Abnormal Psychology, 1967, 72, 193-204.

Chapman, L. J., & Chapman, J. P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. Journal of Abnormal Psychology, 1969, 74, 271-280.

Cohen, R. A. Accuracy and confidence level in clinical inference as a function of cue validity, feedback, and specific practice (Doctoral Dissertation, Kent State University, 1975). Dissertation Abstracts International, 1976, 36, 4682B (Order No. 76-4919, 84).

Costello, C. G. Classification and psychopathology. In C. G. Costello (Ed.), Symptoms of psychopathology: A handbook. New York: John Wiley & Sons, 1970.

Cox, G., Costanzo, P. R., & Coie, J. D. A survey instrument for the assessment of popular conceptions of mental illness. Journal of Consulting and Clinical Psychology, 1976, 44, 901-909.

Crocetti, G. M., Spiro, H. R., Lemkau, P. V., & Siassi, I. Multiple models and mental illnesses: A rejoinder to "Failure of a moral enterprise: Attitudes of the public toward mental illness" by T. R. Sarbin and J. C. Mancuso. Journal of Consulting and Clinical Psychology, 1972, 39, 1-5.

Davis, N. J. Labelling theory in deviance research: A critique and reconsideration. Sociological Quarterly, 1972, 13, 447-474.

Del Gaudio, A. C., Stein, L. S., Ansley, M. V., & Carpenter, P. J. Attitudes of therapists varying in community mental health ideology and democratic values. Journal of Consulting and Clinical Psychology, 1976, 44, 646-655.

Diederich, R. C., Messick, S., & Tucker, L. R. A general least squares solution for successive intervals. Psychometrika, 1957, 22, 159-173.

Di Nardo, P. A. Social class and diagnostic suggestion as variables in clinical judgment. Journal of Consulting and Clinical Psychology, 1975, 43, 363-368.

Endler, N. S. Estimating variance components from mean squares for random and mixed effects analysis of variance models. Perceptual and Motor Skills, 1966, 22, 559-570.

Estell, R. J. Context effects in the psychological scaling of the functional psychotic disorders (Doctoral Dissertation, Fordham University, 1975). Dissertation Abstracts International, 1976, 36, 4151-4152B (Order No. 76-4165, 118).

Farber, I. E. Sane and insane: Constructions and misconceptions. Journal of Abnormal Psychology, 1975, 84, 589-620.

Farina, A., & Felner, R. Employment interviewer reactions to former mental patients. Journal of Abnormal Psychology, 1973, 82, 268-272.

Farina, A., Felner, R., & Boudreau, L. Reactions of workers to male and female mental patient job applicants. Journal of Consulting and Clinical Psychology, 1973, 41, 363-372.

Farina, A., & Ring, K. The influence of perceived mental illness on inter-personal relations. Journal of Abnormal Psychology, 1965, 70, 47-51.

Fitzgibbons, D. J., & Shearn, C. R. Concepts of schizophrenia among mental health professionals: A factor-analytic study. Journal of Consulting and Clinical Psychology, 1972, 38, 288-295.

Foley, J. M. The effect of labeling and teacher attitude on peer acceptance (Doctoral Dissertation, Pennsylvania State University, 1976). Dissertation Abstracts International, 1977, 37, 6361A (No. 77-8884, 73).

Foster, G. G. Expectancy and halo effects as a result of artificially induced teacher bias (Doctoral Dissertation, Pennsylvania State University, 1976). Dissertation Abstracts International, 1976, 37, 2738A (No. 76-24, 764, 138).

Goldberg, L. R. The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender-Gestalt Test. Journal of Consulting Psychology, 1959, 23, 25-33.

Goldberg, L. Simple models or simple processes? Some research on clinical judgments. American Psychologist, 1968, 23, 483-496.

Gordon, R. H. Diagnostic compliance in Rorschach interpretation as a function of group member status. Journal of Consulting and Clinical Psychology, 1976, 44, 826-831.

Gove, W. Societial reaction as an explanation of mental illness: An evaluation. American Sociological Review, 1970a, 35, 873-884.

Gove, W. Who is hospitalized: A critical review of some sociological studies of mental illness. Journal of Health and Human Behavior, 1970b, 11, 294-304.

Gove, W. Individual resources and mental hospitalization: A comparison and evaluation of the societal reaction and psychiatric perspectives. American Sociological Review, 1974, 39, 86-100.

Gove, W., & Fain, T. The stigma of mental hospitalization. Archives of General Psychiatry, 1973, 28, 494-500.

Gross, A. E., & Crofton, C. What is good is beautiful. Sociometry, 1977, 40, 85-90.

Harman, H. H. Modern factor analysis (3rd ed.). Chicago: University of Chicago Press, 1976.

Hoffman, H., & Jackson, D. N. Substantive dimensions of psychopathology derived from MMPI content scales and the DPI. Journal of Consulting and Clinical Psychology, 1976, 44, 862.

Hoffman, H., Jackson, D. N., & Skinner, H. A. Dimensions of psychopathology among alcoholic patients. Journal of Studies on Alcohol, 1975, 36, 825-837.

Horst, P. Factor analysis of data matrices. New York: Holt, Rinehart, and Winston, 1965.

Hunt, W. A., Schwartz, M. L., & Walker, R. E. Reliability of clinical judgments as a function of range of pathology. Journal of Abnormal Psychology, 1965, 70, 32-33.

Jackson, D. N. The multidimensional nature of human social perception. Canadian Journal of Behavioral Science, 1969, 1, 227-262.

Jackson, D. N. A model for inferential accuracy. The Canadian Psychologist, 1972, 13, 185-194.

Jackson, D. N., & Carlson, K. A. Convergent and discriminant validation of the differential personality inventory. Journal of Clinical Psychology, 1973, 29, 214-219.

Jackson, D. N., & Messick, S. Differential Personality Inventory. London, Ontario: Authors, 1971.

Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.

Keppel, G. Design and analysis: A researcher's handbook. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Kirk, S. A. The impact of labeling on rejection of the mentally ill: An Experimental study. Journal of Health and Social Behavior, 1974, 15, 108-117.

Kirk, S. A. The psychiatric sick role and rejection. Journal of Nervous and Mental Disease, 1975, 161, 318-325.

Kirk, S. A. Labelling the mentally ill and the attribution of personal traits. Journal of Consulting and Clinical Psychology, 1976, 44, 506.

Langer, E. J., & Abelson, R. P. A patient by any other name...: Clinician group differences in labeling bias. Journal of Consulting and Clinical Psychology, 1974, 42, 4-9.

Lay, C. H., & Jackson, D. N. Analysis of the generality of trait inferential relationships. Journal of Personality and Social Psychology, 1969, 12, 12-21.

Lee, S. D., & Temerlin, M. K. Social class, diagnosis, and prognosis for psychotherapy. Psychotherapy: Theory, Research, and Practice, 1970, 7, 181-185.

Lindel, M. K. Differential effects of cognitive and outcome feedback in improving judgmental accuracy (Doctoral Dissertation, University of Colorado, 1975). Dissertation Abstracts International, 1975, 36, 2525B (No. 75-23, 615, 59).

Loeber, R., & Weisman, R. G. Contingencies of therapist and trainer performance: A review. Psychological Bulletin, 1975, 82, 660-688.

London, M., & Hakel, M. D. Effects of applicant stereotypes, order, and information on interview impressions. Journal of Applied Psychology, 1974, 59, 157-162.

Manis, M., & Moore, J. C. Summarizing controversial messages: Retroactive effects due to subsequent information. Social Psychology, 1978, 41, 62-68.

Messick, S., & Abelson, R. P. The additive constant problem in multidimensional scaling. Psychometrika, 1956, 21, 1-15.

Messick, S., & Jackson, D. N. Judgmental dimensions of psychopathology. Journal of Consulting and Clinical Psychology, 1972, 38, 418-427.

Morf, M., Syrotuik, J., & Krznaric, S. Real data simulation of a two stage statistical diagnostic system. Journal of Consulting and Clinical Psychology, 1977, 45, 822-828.

Murphy, J. M. Psychiatric labelling in cross-cultural perspective. Science, 1976, 191, 1019-1028.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Oskamp, S. Overconfidence in case study judgments. Journal of Consulting Psychology, 1965, 29, 261-265.

Page, S. The elusive character of psychiatric stigma. Canada's Mental Health, 1974, 22, 15-17.

Page, S. Effects of the mental illness label in attempts to obtain accomodation. Canadian Journal of Behavioral Science, 1977, 9, 85-90.

Paquin, M. J., & Jackson, D. N. Labelling effects in clinical judgments of psychopathology. Journal of Clinical Psychology, 1977, 33, 109-115.

Persons, R. W., & Marks, P. A. Self-disclosure with recidivists: Optimum interviewer-interviewee matching. Journal of Abnormal Psychology, 1970, 76, 387-391.

Reade, W. K., & Wertheimer, M. A bias in the diagnosis of schizophrenia. Journal of Consulting and Clinical Psychology, 1976, 44, 878.

Reed, P. L. Assessing inferential accuracy in clinical judgment and person perception. (Doctoral Dissertation, University of Western Ontario, London, Canada, 1976). Dissertation Abstracts International, 1977, 37, 5333B.

Reed, P. L., & Jackson, D. N. Clinical judgment of psychpathology: A model for inferential accuracy. Journal of Abnormal Psychology, 1975, 84, 475-482.

Ries, J. K. Public acceptance of the disease concept of alcoholism. Journal of Health and Social Behavior, 1977, 18, 338-344.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 6), N.Y.: Academic Press, 1972(a).

Rosenberg, S., & Sedlak, A. Structural representations of perceived personality trait relationships. In A. K. Rooney, R. N. Shepard, & S. B. Nerlove (Eds.) Multidimensional scaling: Theory and application in the behavioral sciences (Vol. 2), N.Y.: Seminar Press, 1972(b).

Rosenhan, D. L. On being sane in insane places. Science, 1973, 179, 250-258.

Rosenhan, D. L. The contextual nature of psychiatric diagnosis. Journal of Abnormal Psychology, 1975, 84, 462-474.

Sarbin, T. R. On the futility of the proposition that some people be labeled "mentally ill". Journal of Consulting Psychology, 1967, 31, 447-453.

Sarbin, T. R., & Mancuso, J. C. Paradigms and moral judgments: Improper conduct is not a disease. Journal of Consulting and Clinical Psychology, 1972, 39, 6-8.

Schacht, T., & Nathan, P. E. But is it good for the psychologists? Appraisal and status of DSM-III. American Psychologist, 1977, 32, 1017-1025.

Schectman, F. Provocative issues in psychiatric diagnosis: A dialogue. Bulletin of the Menninger Clinic, 1976, 40, 435-458.

Scheff, T. The labeling theory of mental illness. American Sociological Review, 1974, 39, 444-452.

Scheff, T. J. (Ed.). Labeling Madness. Englewood Cliffs, N.J.: Prentice Hall, 1975.

Scheid, A. B. Clients' perception of the counselor: The influence of counselor introduction and behavior. Journal of Counseling Psychology, 1976, 23, 503-508.

Simmons, R. E. Bias in the judgment of psychopathology (Doctoral Dissertation, Catholic University of America, 1975). Dissertation Abstracts International, 1975, 36, 1460-1461B (No. 75-19, 522, 52).

Skinner, H. A., Jackson, D. N., & Hoffman, H. Alcoholic personality types: Identification and correlates. Journal of Abnormal Psychology, 1974, 83, 658-666.

Skinner, H. A., Reed, P. L., & Jackson, D. N. Toward the objective diagnosis of psychopathology: Generalizability of modal personality profiles. Journal of Consulting and Clinical Psychology, 1976, 44, 111-117.

Snyder, C. R. "A patient by any other name" revisited: Maladjustment or attributional locus of problem. Journal of Consulting and Clinical Psychology, 1977, 45, 101-103.

Snyder, C. R., Shenkel, R. J., & Schmidt, A. Effects of role perspective and client psychiatric history on locus of problem. Journal of Consulting and Clinical Psychology, 1976, 44, 467-472.

Spitzer, R. L. On pseudoscience in science, logic in remission and psychiatric diagnosis: A critique of Rosenhan's "On being sane in insane places". Journal of Abnormal Psychology, 1975, 84, 442-452.

Spitzer, R. L., & Fleiss, J. L. A re-analysis of the reliability of psychiatric diagnosis. British Journal of Psychiatry, 1974, 125, 341-347.

Staples, F. R., Sloane, R. B., Whipple, K., Cristol, A. H., & Yorkston, N. Process and outcome in psychotherapy and behavior therapy. Journal of Consulting and Clinical Psychology, 1976, 44, 340-350.

Strasburger, E. L., & Jackson, D. N. Improving accuracy in a clinical judgment task. Journal of Consulting and Clinical Psychology, 1977, 45, 303-309.

Stricker, L. J., Jacobs, P. I., & Kogan, N. Trait interrelations in implicit personality theories and questionnaire data. Journal of Personality and Social Psychology, 1974, 30, 198-207.

Sushinsky, L. W., & Wener, R. Distorting judgments of mental health: Generality of the labeling bias effect. Journal of Nervous and Mental Disease, 1975, 161, 82-89.

Temerlin, M. K. Suggestion effects in psychiatric diagnosis. Journal of Nervous and Mental Disease, 1968, 147, 349-353.

Temerlin, M. K. Diagnostic bias in community mental health. Community Mental Health Journal, 1970, 6, 110-117.

Temerlin, M. K., & Trousdale, W. W. The social psychology of clinical diagnosis. Psychotherapy: Theory, Research, and Practice, 1969, 6, 24-29.

Torgerson, W. Theory and methods of scaling. New York: Wiley, 1958.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Ullman, L. P., & Krasner, L. A psychological approach to abnormal behavior. Englewood Cliffs, N.J.: Prentice Hall, 1969.

Vaughan, G. M., & Corballis, M. C. Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. Psychological Bulletin, 1969, 72, 204-213.

Weinberg, H. The effects of labels and social settings on the evaluation of interviewers presenting normal and disturbed behavior (Doctoral Dissertation, Washington University, 1976). Dissertation Abstracts International, 1977, 37, 4173-4174B (No. 77-4058, 189).

Weiner, B. "On being sane in insane places": A process (attributional) analysis and critique. Journal of Abnormal Psychology, 1975, 84, 433-441.

Wyatt, G. E., Reardon, D. F., & Bass, B. A. The readjustment of black, high-risk adolescents to the community. Journal of Community Psychology, 1977, 5, 72-78.

Yaffe, P. E., & Mancuso, J. C. Effects of therapist behavior on peoples' mental illness judgments. Journal of Consulting and Clinical Psychology, 1977, 45, 84-91.

Zigler, E., & Phillips. Psychiatric diagnosis: A critique. Journal of Abnormal and Social Psychology, 1961, 63, 607-618.

Zubin, J. Classification of the behavior disorders. Annual Review of Psychology, 1967, 18, 373-406.

APPENDIX A

The Instructions, Stimuli, Modal Profile (personality) Descriptions and Rating Materials for Study One.

THE SIMILARITY OF PSYCHOPATHOLOGY CONCEPTS: INSTRUCTIONS

PLEASE READ CAREFULLY

1. In order to accurately complete your judgments of the similarity of the concepts in this study, it is necessary that you become fairly familiar with each of the concepts individually. Therefore, please read the list of concepts on the sheet labelled "Stimuli" carefully, at least twice. Please study the list now before continuing with these instructions.

2. You will have noticed in the list the names of eight individuals. These people are described in terms of the results of actual psychometric test data on the sheet titled "Individual Personality Descriptions". In a moment you will be asked to read each of these descriptions, carefully, twice. On the first reading concentrate on trying to associate the description with the name of the person it depicts. Then read it again, concentrating on the specific information contained within the paragraph description. Now please go ahead and study the Individual Personality Descriptions as instructed.

3. Please turn to the Item Booklet but keep handy the Individual Personality Descriptions and List of Stimuli for quick reference. You may now begin to judge the similarity of each of the stimulus pairs by assigning a number from 1 to 9 to each pair and recording this number in the matching numbered space on the answer sheet. The higher the number the more that pair are judged similar. Thus, the rating number represents your judgment of the extent to which the two concepts involved are similar. Below is an example of someone who judged the stimuli "Jack Cole" and "Schizophrenia, paranoid type" as being very similar. This is indicated by the "8" rating. Your own judgment may be different.

Example 1: Item Booklet #214 Schizophrenia, paranoid type Jack Cole

Answer Sheet #214 8

Example 2: Item Booklet #102 drug addict anxiety neurosis

Answer Sheet #102 2

The second example depicts a person who judged the concepts "drug addict" and "anxiety neurosis" as being very dissimilar, thus the low rating number, i.e., "2" which would appear in space #"102___" on the answer sheet.

For your convenience some possible verbal referrents for the rating numbers are suggested below. This is a guideline only. What is critical is that for any item a "9" rating, for example, means you have judged the item as being more similar than an item which you gave an "8" rating.

1 extremely dissimilar;
2 very dissimilar;
3 moderately dissimilar;
4 slightly dissimilar;
5 Equal/Uncertain
6 slightly similar
7 moderately similar
8 very similar
9 extremely similar

TRY TO MAKE USE OF THE ENTIRE RANGE OF RATING NUMBERS.

First impressions are often correct so do not ponder for more than a few seconds on any one stimulus pair.

4. Please now turn to the ITEM BOOKLET and begin the similarity judgments, by rating all the items on page one then proceed to page two, complete all those items and so on. Note that the items are not numbered sequentially by ones so make sure that the similarity rating you assign to a particular item is marked in the space of the same number on the Answer Sheet. When you have finished page ten of the Item Booklet proceed to the final brief, section for instructions, as outlined on page 11.

Part II: Stimulus Rating Instructions

This final section involves the judgment of each of the stimuli individually. Most of the judgments require, as before, 9 point ratings. However, you are now to judge certain aspects of the stimuli rather than the similarity of stimulus pairs. Please turn now to "Answer Sheet Part II," which is on the back of the "Answer Sheet." You will rate each of the 26 stimuli in terms of the questions below, and record your judgment in the appropriate numbered space. Thus, begin with "(a) former psychiatric patient", answer all the questions, then procede to "(b) alcoholic", answer all the questions again in terms of the stimulus, "alcoholic" and so on until you complete the questions for "(z) Bill Dobson".

The Rating Questions

1. How confident were you about the similarity judgments you made involving this stimulus? Indicate this by placing a number between one (1) and the nine (9) such that the more confident you were the higher the number you put in space number "1 ___" for each stimulus. A low rating signifies lack of confidence or uncertainty about the stimulus, while a high rating such as an 8 or 9 means the opposite - high confidence or certainty. For example, if you felt moderately uncertain about your judgements of the stimulus pairs involving "alcoholic" you might put a 3 or 4 in space number 1 beside "(b) alcoholic".

2. How psychopathological do you think this stimulus is? Again place a number from 1 to 9 in the space number "2 ___" for each stimulus; The greater the psychopathology you preceive the higher the rating. A 1 means normality and 9 means extreme psychopathology.

3. In space number "3 ___", for each stimulus, indicate how likely you think it is that detailed psychological and psychiatric examination would reveal the need for some form of institutionalized care for a person described like this stimulus? Your answer should be a number between 1, meaning it is extremely unlikely, to 9, meaning extremely probable or likely.

4. Out of a random sample of 1,000 people, how many do you think could be described by this term? Your answer should be a number between zero (0) and one thousand (1,000), placed in space number "4 ___" for each stimulus.

5. In space number "5 ___", for each stimulus, indicate what percentage of the general population would view this stimulus (person) in a socially undersirable light (i.e."bad" or "sick" or "mentally ill" or "insane" etc.)? Your answer should be a number between zero (0) and one hundred (100).

Finally, rate each of the stimuli on the following adjective scales and place your answer (number) in the corresponding numbered space. Choose a number, closest to the end of each scale, which best describes the stimulus. For example, if you thought that the stimulus "alcoholic" could be described as being quite passive (rather than active) then you would put a 3 or maybe a 2 in space number "6 ___" for row (b).

6.	passive	1	2	3	4	5	6	7	8	9	active
7.	stable	1	2	3	4	5	6	7	8	9	changeable
8.	good	1	2	3	4	5	6	7	8	9	bad
9.	sane	1	2	3	4	5	6	7	8	9	insane
10.	weak	1	2	3	4	5	6	7	8	9	strong

STIMULI

a. former psychiatric patient	b. alcoholic
c. ex-convict	d. psychiatric hospital patient
e. drug addict	f. insane (person)
g. mentally disturbed (person)	h. nervous (person)
i. schizophrenia, simple type	j. schizophrenia, simple type, in remission*
k. schizophrenia, paranoid type	l. schizophrenia, paranoid type, not presently psychotic
m. anxiety neurosis	n. depressive neurosis
o. manic-depressive illness, depressed type	p. manic-depressive illness, depressed type, mild
q. marital maladjustment	r. alcoholism, episodic excessive drinking
s. Alan Hill	t. Jack Cole
u. John Walker	v. Art Reynolds
w. Jim Anderson	x. Brian Patz
y. Peter Carling	z. Bill Dobson

A brief description, in the form of a personality profile, is found on the next page for each of the last eight stimuli (people). Please refer to your instructions.

* "in remission" applies to "conditions that consist of episodes separated by sympton-free intervals" and can be taken to mean "a temporary lessening of a disease or pain."

Note: The defining DPI scales, and other details for the eight personality profiles are outlined in the following table of the Target Characteristics. N.B., This information was not presented to the judges.

INDIVIDUAL PERSONALITY DESCRIPTIONS

(s) Allan Hill "Alan Hill is a rather unemotional man who seldom cares strongly either way about things. He does however, enjoy topics involving horror or suffering or punishment because they are somehow thrilling to him. Alan also feels that certain people in his life are against him and he feels threatened by them. He would even prefer that they did not talk to other people he knows. This is due to his skepticism about people's motives and sincerity and because he thinks that they are concerned mostly with their own self interests."

(u) John Walker "John Walker is somewhat pre-occupied with his health and has consulted his doctor several times in the past month. John is quite unemotional and seldom takes a strong stand on things. He avoids unpleasant, difficult or exciting topics and becomes anxious when talking or thinking about things he has done, or which have happened to him in the recent past. His friends say he is never saddened or overjoyed by anything."

(t) Jack Cole "Jack Cole has been arrested several times for theft. Usually his crimes have been rather poorly planned and reckless. He says he does not feel guilty about his behavior and often explains his stealing something by simply saying that he wanted it. In interviews Jack frequently mentioned his strong dislike for rules and discipline, and he seldom speaks of friends."

(v) Art Reynolds "Art Reynolds is quite happy with his life. He is married, has two children and a successful business of his own. He finds his work to be interesting and challenging but he is still able to find time for his family and other interests as well. While Art of course, has some problems and disappointments, most of his friends agree that things are going will for him."

(w) Jim Anderson Recently Jim has been feeling very downhearted and "blue". He engages in little activity, and is rarely seen with other people. Jim holds a very bad opinion of himself and generally thinks he is pretty worthless. When his friends talk to him about the way he feels, Jim becomes noticeably uncomfortable and tries to change the subject. In general, they find that Jim is not really aware of any reasons for these feelings and behaviors.

(x) Brian Patz Brian Patz is a person who is prone to undertake risky and reckless actions largely because he is unable to think beyond the present about the consequences of his behavior. Brian finds it difficult to focus his attention on everyday details and his collegues say he is quite absent minded and forgetful.

(y) Peter Carling Peter Carling is an unemotional guy characterized as being disinterested and uninvolved by those who know him. However, he always thinks he is sick and thus is constantly complaining about having pains or bodily dysfunctions. His physician reports that Peter lacks energy and general good health and is prone to misinterpret what he sees and hears. Occaisionally, he even experiences visions or strange voices.

(z) Bill Dobson Bill Dobson is experiencing conflict with his family and thus for quite some time has reported his family relationship as being dissatisfying. He has been accused by his family of often misinterpreting discussions and interactions. Bill sometimes talks to absent objects and people but denies having any unfavorable traits, even common ones, and has very little capacity for self criticism.

Characteristics of the Eight Target Descriptions

	Target Name	Modal Profile[a]	Defining DPI Scales
1	Alan Hill	5 +	Cynicism, Ideas of Persecution Sadism and Shallow Affect
2	Jack Cole	2 +	Desocialization, Impulsivity, Socially Deviant Attitudes, Rebelliousness, Sadism
3	John Walker	1 +	Health Concern, Repression, Shallow Affect
4	Art Reynolds	Normal	None ("Normal" personality)
5	Jim Anderson	3 -	Desocialization, Repression, Self Depreciation, Depression
6	Brian Patz	4 +	Impulsivity, Neurotic Disorganization
7	Peter Carling	6 -	Hypochondriasis, Somatic Complaints, Perceptual Distortion, Shallow Affect
8	Bill Dobson	7 -	Familial Discord, Defensiveness, Perceptual Distortion

[a] This column refers to the empirically identified, profile type which defined the targets; they are derived from a multiprofile analysis of five clinical and three college populations described in Skinner, Reed and Jackson (1976). The profile types used in this study were developed in a study by Reed (1976).

ANSWER SHEET

NOTE: Each column covers one page of the Item Booklet

1 ____	2 ____	3 ____	4 ____	5 ____	6 ____	7 ____	8 ____	9 ____	10 ____
11 ____	12 ____	13 ____	14 ____	15 ____	16 ____	17 ____	18 ____	19 ____	20 ____
21 ____	22 ____	23 ____	24 ____	25 ____	26 ____	27 ____	28 ____	29 ____	30 ____
31 ____	32 ____	33 ____	34 ____	35 ____	36 ____	37 ____	38 ____	39 ____	40 ____
41 ____	42 ____	43 ____	44 ____	45 ____	46 ____	47 ____	48 ____	49 ____	50 ____
51 ____	52 ____	53 ____	54 ____	55 ____	56 ____	57 ____	58 ____	59 ____	60 ____
61 ____	62 ____	63 ____	64 ____	65 ____	66 ____	67 ____	68 ____	69 ____	70 ____
71 ____	72 ____	73 ____	74 ____	75 ____	76 ____	77 ____	78 ____	79 ____	80 ____
81 ____	82 ____	83 ____	84 ____	85 ____	86 ____	87 ____	88 ____	89 ____	90 ____
91 ____	92 ____	93 ____	94 ____	95 ____	96 ____	97 ____	98 ____	99 ____	100 ____
101 ____	102 ____	103 ____	104 ____	105 ____	106 ____	107 ____	108 ____	109 ____	110 ____
111 ____	112 ____	113 ____	114 ____	115 ____	116 ____	117 ____	118 ____	119 ____	120 ____
121 ____	122 ____	123 ____	124 ____	125 ____	126 ____	127 ____	128 ____	129 ____	130 ____
131 ____	132 ____	133 ____	134 ____	135 ____	136 ____	137 ____	138 ____	139 ____	140 ____
141 ____	142 ____	143 ____	144 ____	145 ____	146 ____	147 ____	148 ____	149 ____	150 ____
151 ____	152 ____	153 ____	154 ____	155 ____	156 ____	157 ____	158 ____	159 ____	160 ____
161 ____	162 ____	163 ____	164 ____	165 ____	166 ____	167 ____	168 ____	169 ____	170 ____
171 ____	172 ____	173 ____	174 ____	175 ____	176 ____	177 ____	178 ____	179 ____	180 ____
181 ____	182 ____	183 ____	184 ____	185 ____	186 ____	187 ____	188 ____	189 ____	190 ____
191 ____	192 ____	193 ____	194 ____	195 ____	196 ____	197 ____	198 ____	199 ____	200 ____
201 ____	202 ____	203 ____	204 ____	205 ____	206 ____	207 ____	208 ____	209 ____	210 ____
211 ____	212 ____	213 ____	214 ____	215 ____	216 ____	217 ____	218 ____	219 ____	220 ____
221 ____	222 ____	223 ____	224 ____	225 ____	226 ____	227 ____	228 ____	229 ____	230 ____
231 ____	232 ____	233 ____	234 ____	235 ____	236 ____	237 ____	238 ____	239 ____	240 ____
241 ____	242 ____	243 ____	244 ____	245 ____	246 ____	247 ____	248 ____	249 ____	250 ____
251 ____	252 ____	253 ____	254 ____	255 ____	256 ____	257 ____	258 ____	259 ____	260 ____
261 ____	262 ____	263 ____	264 ____	265 ____	266 ____	267 ____	268 ____	269 ____	270 ____
271 ____	272 ____	273 ____	274 ____	275 ____	276 ____	277 ____	278 ____	279 ____	280 ____
281 ____	282 ____	283 ____	284 ____	285 ____	286 ____	287 ____	288 ____	289 ____	290 ____
291 ____	292 ____	293 ____	294 ____	295 ____	296 ____	297 ____	298 ____	299 ____	300 ____
301 ____	302 ____	303 ____	304 ____	305 ____	306 ____	307 ____	308 ____	309 ____	310 ____
311 ____	312 ____	313 ____	314 ____	315 ____	316 ____	317 ____	318 ____	319 ____	320 ____
321 ____	322 ____	323 ____	324 ____	325 ____					

PLEASE FILL IN NAME HERE: _____

ANSWER SHEET PART II

	STIMULUS	RATING QUESTIONS
(a)	former psychiatric patient	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(b)	alcoholic	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(c)	ex-convict	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(d)	psychiatric hospital patient	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(e)	drug addict	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(f)	insane (person)	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(g)	mentally disturbed (person)	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(h)	nervous (person)	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(i)	schizophrenia, simple type	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(j)	schizophrenia, simple type in remission	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(k)	schizophrenia, paranoid type	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(l)	schizophrenia, paranoid type, not presently psychotic	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(m)	anxiety neurosis	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(n)	depressive neurosis	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(o)	manic-depressive illness, depressed type	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(p)	manic-depressive illness, depressed type, mild	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(q)	marital maladjustment	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(r)	alcoholism, episodic excessive drinking	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(s)	Alan Hill	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(t)	Jack Cole	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(u)	John Walker	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(v)	Art Reynolds	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(w)	Jim Anderson	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(x)	Brian Patz	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(y)	Peter Carling	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___
(z)	Bill Dobson	1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 ___ 8 ___ 9 ___ 10 ___

APPENDIX B

The Instructions, Rating Scales, DPI Item Information, and Sample Assessment Reports for Study Two

The Nature and Accuracy of Mental Illness and Psychopathology Judgments

INSTRUCTIONS

N.B. IT IS ESSENTIAL THAT YOU READ THESE INSTRUCTIONS VERY CAREFULLY.

The purpose of the study is to examine processes in making decisions about various kinds of clinical judgments. Studies of these factors can improve the accuracy of such judgments by enhancing our culturally learned, largely accurate, conceptions about the relationships among various behavior patterns and traits. On the following pages you will find brief descriptions of four individuals, (whose real names shall remain confidential) who have undergone a routine psychological assessment as part of a large company's employee benefits program.

The assessment reports have been prepared by a variety of mental health personnel (e.g., psychiatrists, psychologists, social workers, testing clerks) in a large mental health facility in this area. You will be judging the test summaries of one such "assessor" who has completed an assessment on numerous individuals according to a standardized report format. Thus, the summary contains background information and history, a summary of major traits and findings, and a diagnosis about severity with respect to the frequency of occurrence of the identified behavior patterns.

All identifying and specific information has been removed from or modified on the transcripts; this accounts for the lack of flow of a few parts of the description. Since the descriptions are concise summaries of detailed reports prepared according to the standardized format you are asked to carefully read the descriptions, twice, according to these instructions. Otherwise you may miss vital information.

over...

First, read the summary assessment and associate the main characteristics with the name of the person (target) involved. Then read it again concentrating your attention on the nature of the information contained in it. Then, using the "Answer Sheet" record your judgment about the likelihood that the described person would respond "true" to each of the items in the accompanying "Statement Booklet." This is recorded on the answer sheet by putting a circle around the number which you think best reflects the probability of the described person answering true to that item.

For example, if you felt that the described person would almost certainly respond true to item:

(A) I feel sad most of the time,

then you would circle the number "8" or maybe even "9" on the Answer Sheet as below:

	Extremely Unlikely				Neither likely nor unlikely				Extremely Likely
(A)	1	2	3	4	5	6	7	(8)	9

When making your judgments try to make use of the entire range of these nine rating categories. You may refer back to any of the material at any time. When you have completed the 52 items proceed to the section titled "Part II: Rating Instructions" which follows the last statement item (on the next page).

Answer the rating questions for each person as instructed and enter your responses to these questions on the "Answer Sheet Part II", which appears on the back of the main Answer Sheet (Part I). Make sure your answer to each question is placed in the correct numbered space for each person, as instructed.

When you have completed all rating questions proceed to the assessment report of the next described person following the same procedure (for Parts I and II), and so on, until you have completed all four assessment reports.

Part II: Rating Instructions

This final section involves general judgments about the described (target) person. Most of the judgments require, as before, 9 point ratings. Please turn now to "Answer Sheet (Part II)." Please rate the person in terms of the questions, below (53 to 69) and record your judgment in the appropriate numbered space on the answer sheet. Then proceed to the person described in the next assessment report and complete the judgments as instructed.

The Rating Questions

53. How confident were you about the 52 judgments you made involving this person? Indicate this by placing a number between one (1) and nine (9) such that the more confident you were the higher the number you put in space number "53. ___", on the answer sheet (Part II). A low rating signifies lack of confidence or uncertainty, while a high rating such as an 8 or 9 means the opposite = high confidence or certainty. For example, if you felt moderately uncertain about your judgments of this person you might put a 3 or 4 in space number 52.

54. How psychopathological do you think this person is? Again place a number from 1 to 9 in the space number "54. ___" of the answer sheet. The greater the psychopathology you perceive the higher the rating. A 1 means normality and 9 means extreme psychopathology.

55. In space number "55. ___", indicate how likely you think it is that detailed psychological and psychiatric examination would reveal the need for some form of institutionalized care for a person described like this? Your answer should be a number between 1, meaning it is extremely unlikely, to 9, meaning extremely probable or likely that this person requires institutionalized care.

56. Out of a random sample of 1,000 people, how many do you think could be described like, or, are similar to this person? Your answer should be a number between zero (0) and one thousand (1,000).

57. In space number "57. ___", indicate what percentage of the general population would view this person in a socially undesirable light (i.e., "bad" or "sick" or "mentally ill" or "insane" etc.)? Your answer should be a number between zero (0) and one hundred (100).

58. What is the likelihood that this person would derive benefit from appropriate therapy? Place a number from 1 to 9 in space "52. ___" for each person, such that a 1 means extremely unlikely and a 9 means extremely likely.

59. Should this person's problems be alleviated how likely are they to recur in the next five years? Again place a number from 1 (extremely unlikely) to 9 (extremely likely) in the proper space.

over...

Finally, rate each of the people on the following adjective scales and place your answer in the corresponding numbered space on the answer sheet. (Do not mark answers on this sheet.) Choose a number, closest to the end of each scale, which best describes that person. For example, if you thought that the person could be described as being quite passive (rather than active) then you record a 3 or maybe a 2 in item number "60. ___" for that person. If you thought the person was very "changeable" you would place a 9 in space 61 on the answer sheet, etc.

60.	passive	1	2	3	4	5	6	7	8	9	active
61.	stable	1	2	3	4	5	6	7	8	9	changeable
62.	good	1	2	3	4	5	6	7	8	9	bad
63.	sane	1	2	3	4	5	6	7	8	9	insane
64.	weak	1	2	3	4	5	6	7	8	9	strong
65.	mature	1	2	3	4	5	6	7	8	9	youthful
66.	sensitive	1	2	3	4	5	6	7	8	9	insensitive
67.	excitable	1	2	3	4	5	6	7	8	9	calm
68.	cautious	1	2	3	4	5	6	7	8	9	rash
69.	warm	1	2	3	4	5	6	7	8	9	cold

ANSWER SHEET (PART I)

Your Name _____ Target Person's Name _____
(Please print)

	Extremely Unlikely				Neither Likely Nor Unlikely				Extremely Likely			Extremely Unlikely				Neither Likely Nor Unlikely				Extremely Likely	
1.	1	2	3	4	5	6	7	8	9		27.	1	2	3	4	5	6	7	8	9	
2.	1	2	3	4	5	6	7	8	9		28.	1	2	3	4	5	6	7	8	9	
3.	1	2	3	4	5	6	7	8	9		29.	1	2	3	4	5	6	7	8	9	
4.	1	2	3	4	5	6	7	8	9		30.	1	2	3	4	5	6	7	8	9	
5.	1	2	3	4	5	6	7	8	9		31.	1	2	3	4	5	6	7	8	9	
6.	1	2	3	4	5	6	7	8	9		32.	1	2	3	4	5	6	7	8	9	
7.	1	2	3	4	5	6	7	8	9		33.	1	2	3	4	5	6	7	8	9	
8.	1	2	3	4	5	6	7	8	9		34.	1	2	3	4	5	6	7	8	9	
9.	1	2	3	4	5	6	7	8	9		35.	1	2	3	4	5	6	7	8	9	
10.	1	2	3	4	5	6	7	8	9		36.	1	2	3	4	5	6	7	8	9	
11.	1	2	3	4	5	6	7	8	9		37.	1	2	3	4	5	6	7	8	9	
12.	1	2	3	4	5	6	7	8	9		38.	1	2	3	4	5	6	7	8	9	
13.	1	2	3	4	5	6	7	8	9		39.	1	2	3	4	5	6	7	8	9	
14.	1	2	3	4	5	6	7	8	9		40.	1	2	3	4	5	6	7	8	9	
15.	1	2	3	4	5	6	7	8	9		41.	1	2	3	4	5	6	7	8	9	
16.	1	2	3	4	5	6	7	8	9		42.	1	2	3	4	5	6	7	8	9	
17.	1	2	3	4	5	6	7	8	9		43.	1	2	3	4	5	6	7	8	9	
18.	1	2	3	4	5	6	7	8	9		44.	1	2	3	4	5	6	7	8	9	
19.	1	2	3	4	5	6	7	8	9		45.	1	2	3	4	5	6	7	8	9	
20.	1	2	3	4	5	6	7	8	9		46.	1	2	3	4	5	6	7	8	9	
21.	1	2	3	4	5	6	7	8	9		47.	1	2	3	4	5	6	7	8	9	
22.	1	2	3	4	5	6	7	8	9		48.	1	2	3	4	5	6	7	8	9	
23.	1	2	3	4	5	6	7	8	9		49.	1	2	3	4	5	6	7	8	9	
24.	1	2	3	4	5	6	7	8	9		50.	1	2	3	4	5	6	7	8	9	
25.	1	2	3	4	5	6	7	8	9		51.	1	2	3	4	5	6	7	8	9	
26.	1	2	3	4	5	6	7	8	9		52.	1	2	3	4	5	6	7	8	9	

over...

ANSWER SHEET (Part II)

Question #

53. ____
54. ____
55. ____
57. ____
58. ____
59. ____
60. ____
61. ____
62. ____
63. ____
64. ____
65. ____
66. ____
67. ____
68. ____
69. ____

Description of DPI Booklet Items

Booklet Number	Item	DPI Scale	DPI Number	Scoring Key	Endorsement[a] Proportion
1.	I often wake up during the night.	Insomnia	55	T	.24
2.	I am not afraid that anyone will steal my ideas.	Broodiness	299	F	.22
3.	I am as happy as my friends.	Depression	247	F	.18
4.	There is good reason to believe that "there's a sucker born every minute."	Cynicism	138	T	.39
5.	I like to do the opposite of what other people do.	Rebelliousness	73	T	.27
6.	Some movies cause me to become quite emotional.	Shallow Affect	374	F	.16
7.	I often tell others that I do not like a person.	Hostility	280	T	.30
8.	The way I feel about my life changes quite often.	Mood Fluctuation	420	T	.36
9.	I believe most people do not lie on purpose.	Cynicism	273	F	.37
10.	Others always seem to enjoy life more than I.	Depression	328	T	.17
11.	When I get pains, I can't tell other people what they are like.	Hypochondriasis	38	T	.13
12.	I worry very little about my health.	Health Concern	414	F	.24
13.	I'd like to meet a famous criminal.	Socially Deviant Attitudes	429	T	.28
14.	When I do things, I usually do them quite well.	Self Depreciation	400	F	.11
15.	Even if a criminal is successful, I don't think he is worthy of praise.	Socially Deviant Attitudes	24	F	.41

Description of DPI Booklet Items

Booklet Number	Item	DPI Scale	DPI Number	Scoring Key	Endorsement Proportion
16.	Little things often bother me.	Irritability	392	T	.41
17.	Sometimes my surroundings appear to change so that I am in a strange place.	Feelings of Unreality	62	T	.20
18.	I like to work with a group of people rather than alone.	Disocialization	329	F	.40
19.	When people whisper, I feel they might be talking about me.	Broodiness	326	T	.23
20.	I have a good night's sleep just about every night.	Insomnia	1	F	.18
21.	I seldom get a "splitting headache".	Headache Proneness	82	F	.16
22.	I don't either like or dislike people.	Shallow Affect	23	T	.18
23.	Rarely, if ever, do I do foolish things without thinking.	Impulsivity	175	F	.36
24.	Whenever I am worried about something, I get a headache.	Headache Proneness	28	T	.13
25.	Familiar things never seem "foggy" or far away to me.	Feelings of Unreality	89	F	.21
26.	I cannot remember how to do things which used to be easy for me.	Disorganization of Thinking	60	T	.11
27.	I seldom make much effort to meet new people.	Desocialization	86	T	.39
28.	I feel free to say what I think at home without fear of getting into an argument.	Familial Discord	277	F	.37
29.	I don't very often feel like "telling someone off".	Hostility	37	F	.28

Description of DPI Booklet Items

Booklet Number	Item	DPI Scale	DPI Number	Scoring Key	Endorsement Proportion
30.	I like to think about difficult questions even if I can't find answers for them.	Repression	398	F	.27
31.	I seldom catch colds.	Somatic Complaints	376	F	.23
32.	Many times I do things without thinking.	Impulsivity	256	T	.39
33.	Rarely, if ever, do the habits of other people bother me.	Irritability	365	F	.45
34.	I get short of breath easily.	Somatic Complaints	25	T	.25
35.	I start to feel scared when I think about the things that worry me.	Panic Reaction	395	T	.30
36.	The way I feel about people does not change very much.	Mood Fluctuation	123	F	.42
37.	When I have set a plan for my time, I do not have any trouble keeping to it.	Neurotic Disorganization	394	F	.34
38.	I can hear a voice warn me when I do anything evil.	Perceptual Distortion	423	T	.12
39.	I feel that I should apologize for most of the things I have done.	Self Depreciation	319	T	.11
40.	I sometimes enjoy teasing animals.	Sadism	237	T	.20
41.	I do not get upset with people who startle me.	Panic Reaction	206	F	.30
42.	I cannot hear thought waves of people around me.	Perceptual Distortion	342	F	.14
43.	I often forget to do things that I should do.	Neurotic Disorganization	421	T	.32

Description of DPI Booklet Items

Booklet Number	Item	DPI Scale	DPI Number	Scoring Key	Endorsement Proportion
44.	At night I rarely think over what has happened to me during the day.	Repression	263	T	.29
45.	People considered me obedient as a child.	Rebelliousness	208	F	.27
46.	No one spreads rumors to turn my friends against me.	Ideas of Persecution	255	F	.12
47.	I don't think there is anything wrong with my health.	Hypochondriasis	173	F	.13
48.	I like to have X-rays taken in case something is wrong which wouldn't show up in a regular examination.	Health Concern	117	T	.34
49.	My parents did not seem to understand me.	Familial Discord	142	T	.34
50.	I often feel that someone is trying to make my life difficult and unpleasant.	Ideas of Persecution	12	T	.09
51.	I can easily find my way around when I am left alone.	Disorganization of Thinking	33	F	.11
52.	Accidents are never thrilling to me.	Sadism	372	F	.20

[a] Proportion of subjects responding to the item in the keyed direction, based on a sample of 179 college students.

Note. The Differential Personality Inventory (DPI) is copyrighted 1970 by Douglas N. Jackson and Samuel J. Messick. Reprinted by permission.

Assessment Report

NAME: Alan B. Hill

D.O.B.: January 14, 1949

History: Normal Medical Background

Chart #: 77-0413-3

Date of Testing: September 10, 1976

Tests Administered: Differential Personality Inventory

Tested by: Marjorie Johnson, M.S.W.
Social Work Department

Reason for Referral: Routine assessment, employment services order

Test Summary

In conclusion, Mr. Hill is a rather unemotional man. He seldom cares strongly either way about things. He does however, enjoy topics involving horror or suffering or punishment because they are somehow thrilling to him. Alan also feels that certain people in his life are against him and he feels threatened by them. He would even prefer that they did not talk to other people he knows. This is due to his skepticism about people's motives and sincerity and because he thinks that they are concerned mostly with their own self interests.

The above pattern of traits represents only a minimal likelihood of psychopathology severe enough to be incapacitating. While some uncertainty exists, as always, this person can be diagnosed as only rarely experiences these difficulties.

Marjorie Johnson, M.S.W.
Social Work Department

Assessment Report

NAME: John L. Walker

D.O.B.: April 26, 1952

History: Former psychiatric patient, St. Thomas, Ontario

Chart #: 77-0401-3

Date of Testing: August 19, 1976

Tests Administered: Differential Personality Inventory

Tested by: Janis Clark, Secretary-Clerk Psychology Department

Reason for Referral: Routine assessment, Employment Service Order

Test Summary

To review, Mr. Walker is somewhat pre-occupied with his health, as he was while a former psychiatric patient with us, and has consulted his doctor several times in the past month. John is quite unemotional and seldom takes a strong stand on things. He avoids unpleasant, difficult or exciting topics and becomes anxious when talking or thinking about things he has done, or which have happened to him in the recent past. His friends say he is never saddened or overjoyed by anything.

These problems are not likely to result in any seriously disabling disorders for John. It is particularly important to beware that while those around him may not readily understand it, this trait pettern merely reflects a mild, rarely occurring condition.

Janis Clark, Secretary-Clerk
Psychology Department

Assessment Report

NAME: Jim W. Anderson

D.O.B.: August 1, 1951

History: Former psychiatric patient, Goderich, Ontario

Chart #: 77-0393-3

Date of Testing: August 15, 1976

Tests Administered: Differential Personality Inventory

Tested By: Marjorie Johnson, M.S.W. Social Work Department

Reason for Referral: Routine assessment employment services order

Test Summary

Mr. Anderson, formerly a psychiatric patient, recently has been feeling very downnearted and "blue". He engages in little activity, and is rarely seen with other people. Jim holds a very bad opinion of himself and generally thinks he is pretty worthless. When his friends talk to him about the way he feels, Jim becomes noticeably uncomfortable and tries to change the subject. In general, they find that Jim is not really aware of any reasons for these feelings and behaviors.

Overall, Mr. Anderson reflects a high chance of serious psychopathology. This is associated with probable, undesirable effects for him personally, in a psychosocial sense. Since a unidimensional diagnosis is required it should be noted that he is probably a severe case in terms of the high frequency of these symptoms.

Marjorie Johnson, M.S.W.
Social Work Department

Assessment Report

NAME: Jim W. Anderson

D.O.B.: August 1, 1951

History: Normal Medical Background

Chart #: 77-0393-3

Date of Testing: August 15, 1976

Tests Administered: Differential Personality Inventory

Tested By: G. J. Goddard, Chief Psychiatrist Psychiatric Services

Reason for Referral: Routine assessment, employment services order

Test Summary

Mr. Anderson, recently has been feeling very downhearted and "blue". He engages in little activity, and is rarely seen with other people. Jim holds a very bad opinion of himself and generally thinks he is pretty worthless. When his friends talk to him about the way he feels, Jim becomes noticeably uncomfortable and tries to change the subject. In general, they find that Jim is not really aware of any reasons for these feelings and behaviors.

Overall, Mr. Anderson reflects a high chance of serious psychopathology. This is associated with probable, undesirable effects for him personally, in a psychosocial sense. Since a unidimensional diagnosis is required it should be noted that he is probably a severe case in terms of the high frequency of these symptoms.

G. J. Goddard,
Chief, Psychiatric Services

APPENDIX C

Representative samples of the Instructions
and Clinical Assessment Reports for
Study Three (Chapter Four).

The Nature of Mental Illness and Psychopathology Judgments (II)

INSTRUCTIONS

N.B. IT IS ESSENTIAL THAT YOU READ THESE INSTRUCTIONS VERY CAREFULLY.

INTRODUCTION

The purpose of the study is to examine processes in making decisions about various kinds of clinical judgments. Studies of these factors can improve the accuracy of such judgments by enhancing our culturally learned, largely accurate, conceptions about the relationships among various behavior patterns and traits. On the following pages you will find brief descriptions of four individuals, (whose real names shall remain confidential) who have undergone a routine psychological assessment as part of a large company's employee benefits program.

All identifying and specific information has been removed from or modified on the assessment reports. Since the descriptions are concise summaries of detailed reports prepared according to a standardized format you are asked to carefully read the reports twice, according to these instructions. Otherwise you may miss vital information.

EVALUATING THE REPORTS

(1) First, read the summary assessment and associate the main characteristics with the name of the person (testee) involved. (2) Then read it again concentrating your attention on the nature of the information contained in it. Please focus your attention upon the diagnostic categories (e.g., "paranoid schizophrenic") and severity evaluations (e.g., "high chance of disorder") contained within the report. These summary concepts reflect multiple sources of information within brief, easy to recall words or phrases and thus you should rely heavily on them in your judgments. On the assessment reports these are most like the information segments

which are labelled A, B, C, X, Y, Z. Thus, when answering questions about the described person (testee) your judgments should be most influenced by the A-C and X-Z segments. The letters for the respective segments appear, on the left margin of the report, in parentheses. (3) Once you have carefully studied the report turn to the "Item Booklet" and the "Answer Sheet (Part I)". Using the "Answer Sheet" record your judgment about the likelihood that the described person would respond "true" to each of the items in the "Statement Booklet," based on the vital information contained in the Assessment Report. This is recorded on the answer sheet by putting a circle around the number which you think best reflects the probability of the described person answering true to that item. (4) After rating each item indicate the single most influential information segment involved in your decision about that item. You may only choose one segment.

For example, if you felt that the described person would almost certainly respond true to item:

(1) I feel sad most of the time,

then you would circle the number "8" or maybe even "9" on the Answer Sheet as below: in this case the judge felt that information segment "B" most influenced his rating for item 1.

	Extremely Unlikely				Neither likely nor unlikely				Extremely Likely	Information Segment
(1)	1	2	3	4	5	6	7	(8)	9	B

When making your judgments try to make use of the entire range of these nine rating categories. You may refer back to any of the material at any time. When you have completed the 52 items proceed to the section titled "Part II: Rating instructions" which follows the last item in the Statement Booklet.

Assessment Report

NAME: Peter C. CARLING ; D.O.B.: June 19, 1950

Test administered: Differential Personality Inventory (DPI) Date: Sept. 14, 1976

(A) Reason for Referral: Employer request ; Referral Chart # 77-8-425

(B) History (list details if any): Patient, Goderich Psychiatric Hospital, released May, 1970.

(C) Current Diagnoses: simple schizophrenia

DPI - Test Summary

(D) Mr. Carling is an unemotional man characterized as

(E) being disinterested and univolved by those who know him.

(F) However, he always thinks he is sick and thus is constantly complaining about having pains or bodily dysfunctions.

(G) His physician reports that Peter lacks energy and general good health,

(H) and is prone to misinterpret what he sees and hears.

(I) Occasionally, he even experiences visions or strange voices.

Case Conference Summary

(X) The above pattern of traits reflecting simple schizophrenia represents a moderate probability of serious psychopathology, and

(Y) may result in some disturbing consequences for Mr. Carling.

(Z) While a singular diagnosis is difficult, the above pattern is interesting in that he may suffer from these symptoms with some frequency.

M. J. Johnson, M.S.W.
Social Services Dept.

Assessment Report

NAME: Jack R. COLE ; D.O.B.: Feb. 14, 1950

Test administered: Differential Personality Inventory (DPI) Date: July 26, 1976

(A) Reason for Referral: Employer request; Referral Chart # 77-8-396

(B) History (list details if any): Patient, London Psychiatric Hospital, released July, 1970

(C) Current Diagnoses: ex-convict

DPI Test Summary

(D) Mr. Cole has been arrested several times for theft.

(E) Usually his plans are poorly planned and reckless (this appears to apply to his crimes also).

(F) He says he does not feel guilty about his behavior and

(G) often explains his stealing by simply saying that he wanted it.

(H) In interviews and the test responses he indicates a strong dislike for rules and discipline and,

(I) seldom speaks, nor indicates any need or desire for friends.

Case Conference Summary

(X) Overall, and particularly as an ex-convict Mr. Cole reflects a moderate chance of serious psychopathology.

(Y) This is associated with possible, undesirable effects for him, in a psychosocial sense.

(Z) Since a unidimensional diagnosis (of severity) is required it should be noted that he may be a serious case in terms of the frequency of these symptoms.

M. J. Johnson, M.S.W.
Social Services Dept.

Assessment Report

NAME: John L. WALKER ; D.O.B.: Sept. 2, 1949

Test administered: Differential Personality Inventory (DPI) Date: Aug. 4, 1976

(A) Reason for Referral: Employer request ; Referral Chart # 77-8-431

(B) History (list details if any): Patient, London Psychiatric Hospital, released June, 1971.

(C) Current Diagnoses: ex-convict

DPI Test Summary

(D) Mr. Walker is often pre-occupied with his health,

(E) and has consulted his doctor several times in the past month.

(F) He is quite unemotional and seldom takes a strong stand on things.

(G) He avoids unpleasant, difficult or exciting topics and

(H) becomes anxious when talking or thinking about things he has done, or which have happened to him in the recent past.

(I) His friends confirm that he is never extremely saddened or overjoyed by anything.

Case Conference Summary

(X) These problems, and those associated with his ex-convict status are somewhat likely to result in a seriously disabling disorder for Mr. Walker.

(Y) It is particularly important to recognize that while those around him may not be fully aware of these problems,

(Z) this symptomatology is a fairly frequent condition.

M. J. Johnson, M.S.W.
Social Services Dept.

Assessment Report

NAME: Art W. REYNOLDS ; D.O.B.: May 4, 1948

Test administered: Differential Personality Inventory (DPI) Date: Sept. 9, 1976

(A) Reason for Referral: Employer request ; Referral Chart # 77-8-416

(B) History (list details if any): Patient, St. Thomas Psychiatric Hospital, released August, 1969.

(C) Current Diagnoses: Schizophrenia, simple type

DPI Test Summary

(D) Mr. Reynolds is currently quite happy with his life.

(E) He is married, has two children and a successful business of his own.

(F) He finds his work to be interesting and challenging but,

(G) he still finds time for his family and other interests as well.

(H) While Art of course, has some problems and disappointments,

(I) most of his friends agree, according to Art, that things are going well for him.

Case Conference Summary

(X) In view of the conference's speculation regarding simple schizophrenia it may be advisable to continue monitoring Mr. Reynolds since,

(Y) there is some chance that stressful circumstances will precipitate further symptoms.

(Z) While there is no test based diagnosis as a basis, Mr. Reynolds's repressed tendencies are noteworthy because they appear to occur with some regularity.

M. J. Johnson, M.S.W.
Social Services Dept.

APPENDIX D

Item Selection Information and Description of the Differential Personality Inventory

The trait rating items (see Appendix B) upon which the subjects made their predictions are taken from the Differential Personality Inventory (DPI) of Jackson and Messick (1971). The DPI was designed to assess a number of facets of psychopathology and emphasized the suppression of response styles, substantive generalizability, and convergent and discriminant validity. The test consists of 26 sixteen-item content scales balanced in terms of the number of true and false keyed items. The DPI also incorporates two scales, "Defensiveness" and "Infrequency" which were designed primarily as test taking validity scales.

> "Perhaps the most important departures in the construction of the DPI from the tradition of empirical item selection were in the use of substantively defined, distinctive item pools prepared for each scale and the use of a sequential strategy of item selection," (Jackson & Carlson, 1973, p. 214) concerned with

construct validity. As support for this approach, Jackson and Carlson (1973) present data on the reliability and convergent and discriminant validity of the DPI. The 52 items upon which the subjects made their trait judgments were selected from 26 of the 28 DPI scales. Items from the Defensiveness and Infrequency scales were not included. The items were selected according to the following criteria: (1) One true and one false keyed item from each scale. This reduced the probability of acquiescent responding and related response styles, (2) items with a moderate frequency of endorsement or "p" value were selected in order to minimize social desirability and the effects of extreme information value. The "p" value of the selected items ranged between .09 and .48 with a mean of .26. Furthermore, all items were within half a standard deviation of the mean "p" value of the scale from which they were selected. (3) Only items with reliable "p" values were selected, i.e., those items which manifested a close association for two independent samples of "p" values. The actual items and their corresponding scales were presented in Appendix B.